```
   1 TGCCCGCTGC CCGCCCGCAG TTCCCGGCCC CGCTGGCCCC AGTCATGGCG
  51 AAGCAGTACG ATGTGCTGTT CCGGCTGCTG CTGATCGGGG ACTCCGGGGT
 101 GGGCAAGACC TGCCTGCTGT GCCGCTTCAC CGACAACGAG TTCCACTCCT
 151 CGCACATCTC CACCATCGGT GTTGACTTTA AGATGAAGAC CATAGAGGTA
 201 GACGGCATCA AAGTGCGGAT ACAGATCTGG GACACTGCAG GGCAGGAGAG
 251 ATACCAGACC ATCACAAAGC AGTACTATCG GCGGGCCCAG GGGATATTTT
 301 TGGTCTATGA CATTAGCAGC GAGCGCTCTT ACCAGCACAT CATGAAGTGG
 351 GTCAGTGACG TGGATGAGTA CGCACCAGAA GGCGTCCAGA AGATCCTTAT
 401 TGGGAATAAG GCTGATGAGG AGCAGAAACG GCAGGTGGGA AGAGAGCAAG
 451 GGCAGCAGCT GGCGAAGGAG TATGGCATGG ACTTCTATGA AACAAGTGCC
 501 TGCACCAACC TCAACATTAA AGAGTCATTC ACGCGTCTGA CAGAGCTGGT
 551 GCTGCAGGCC CATAGGAAGG AGCTGGAAGG CCTCCGGATG CGTGCCAGCA
 601 ATGAGTTGGC ACTGGCAGAG CTGGAGGAGG AGGAGGGCAA ACCCGAGGGC
 651 CCAGCGAACT CTTCGAAAAC CTGCTGGTGC TGAGTCCTGT GTGGGGCACC
 701 CCACACGACA CCCCTCTTCC CTCAGGAGGC CCGTGGGCAG ACAGGGGAGC
 751 CGGGGCTTTG CCCTGCTGCT GTCCTCTCGT GTGATGACCC TATTGAGTAT
 801 CAGTAGCCAC TACTCCCCCT GCCTGGCCCT GAGAGCGGCT CTGCTGTCAT
 851 CTCAAGCAGC CCCTGTCCCC AGCCCGTCCA CCCTGGAGTG GTCTTCTTCA
 901 GCCTGTTTCC CCAGCCACAG GCCTGCTACG ACCCCCACGA TGTGCCGCAA
 951 GCACTGTCTC ACCATCCCGC ACCCACCAGA CAACAGCCAG GGCTGGAGTC
1001 CAGGCCACTT TCAGCTGCTC CTTTCTCCGT GCATCGTGTC TCTTCTCTGC
1051 TTTTTCTCTC TTCCCCCACT TCTCTTTCTC TGACCCCTCC CCTCCGGTGC
1101 GTTTCGTATC AAAGCTCCTC AAACCCCGTC CCCGTGTGT CCTGCTGTGT
1151 GCAGCTCGCT CTTTCCTTCC TTCCTAAGCT ATCCAAGGGG ATGGACCCAG
1201 GCTCGTGGGG AGGTTCCACC CTTGGATCCA GGAAGAACCC TCCACCCTGC
1251 CTCGTGGGTG GGCCAAAGGC TACAGGGTGC TTCTTCCTCT TCCCCCACCC
1301 CCACTGTCCC TCATGTGCCA TGGGCCTGCC TCCCCAGTGA CCTGCGAAAG
1351 TGGAGCATCG AGGTAGGAGG GAAACAGCAA CCGGGGAGTC CTCGAGCCTG
1401 GGGCTGCCCT ACCTCTACCC ATTCCCCGAC CAGAGCTTTG CCCTTGCTTG
1451 GCTGCCCGCC TGCCTCTTTG GGGAACTGAG CTCAGAGGCA GGTGCTTCAG
1501 AGAAGGAAAC AAAATGAGGG GTGGCAGGGA TAAAAAGTCA CCTCCATTCT
1551 CTACCTCCCA TGCAGCATGA ACACAATTTC TCTCCACCTG GCTCCCAAAT
1601 TTAAAGATGT GGACCAAGGC CTGTGGGTAC TCCAGGGGCA AGGAGAGCCC
1651 TGGGGTCAGT GACACTGTCA GGCCAACCAT GCACTCCACA AAGGGGAGCA
1701 TTTGGAAATG AAGGACTAGC TCCTATGTAT CAGGTTAAGA GCAAGGGAGA
1751 GCTGGCCAGG GACAGCAGTT TGCACAGCAG AGGGGAATGT AGCAACAGCA
1801 GGGCCTCCTA GGCCCCATCT TCCATTTCTT AGGTAAGAAG AGCATTTCCT
1851 CAGACTCCCA GGCGGAGGAC TGAGCCTAGC CTTCAGCAAC CAAGGTTCTC
1901 CTGGGACCCA AAGTTTATGG GAGAAGGGCA AAGACTTCAT GGGAAGAGAG
1951 AAGGAAGGCC CTGGGTAGAA ACGCTTGGTG CTGTTCTCTT TGGCCTTTAA
2001 GACAAAGCGC TCATCTTGCC CTCTACCTCC TGATAGGCTT GAGGGTTTGC
2051 CAACCACACT GTGGCTACAG GTGGAGGGAA GAGGACTCCT TCCTCCAGAG
2101 TGCTATGTTC AGGAAGTTTC TTTAACCCCA TATGGCCCAA GAGTAGCTCG
2151 TAGGAGGCCC TTTAAAGACG GAACAAGTAA TTTACCAGTT CTACTGGGGT
2201 TCCTGCCCAC CGTCCCAAGG TGGGCGAGGC CTAGGAAGAG GGTCATTCTT
2251 AAGCCACACA TTAGCTGCAC TGCGTGGCTG CAGCCAAAAC AAAGAACTGG
2301 GTGTTGAGTA TTCATCAACT AAGAACCAAA ATCCAGGGCA CTCATATGTG
2351 AAGGATAAGA ACCTCACTTC CTTACTCCTC CAAAAAGAAG TGGGGAAAGA
2401 ACCATCAAAC CTTTCCTCCT GACTTACCAA ACCAGGAAAA CAGCAGGAGA
2451 GGGTGGCTCA GGACTTAGGG ACAGGGTATA GCTTAGATGG TGGAAAGCAA
2501 AGGAGAGCAG GAAGTTGTAA ATCACTGGCT AATGAGAAAA GGAGACAGCT
2551 AACTCTAGGA TGAAGCTGTG ACTAGGCTGG AGTTGCTTCC TTGAAGATGG
2601 GACTCCTTGG GTATCAAGAC CTATGCCACA TCACACTGGG GCTAGGGAAG
2651 TAGGTGATGC CAGCCCTCAA GTCTGTCTTC AGCCAGGGAC TTGAGAAGTT
2701 ATATTGGGCA GTGGCTCCAA TCTGTGGACC AGTATTTCAG CTTTCCCTGA
2751 AGATCAGGCA GGGTGCCATT CATTGTCTTT CTCTCCTAGC CCCCTCAGGA
2801 AAGAAGGACT ATATTTGTAC TGTACCCTAG GGGTTCTGGA AGGGAAAACA
2851 TGGAATCAGG ATTCTATAGA CTGATAGGCC CTATCCACAA GGGCCATGAC
2901 TGGGAAAAGG TATGGGAGCA GAAGGAGAAT TGGGATTTTA GGGTGCAGCT
2951 ACGCTCACCC TAAACTTTTG GTGGCCTGGG GCATGTCTTG AGGCCCAGAC
3001 TGTTAAGCAG GCTCTGCTGG CCTGTTTACT CGTCACCACC TCTGCACCTG
3051 CTGTCTTGAG ACTCCATCCA GCCCCAGGCA CGCCACCTGC TCCTGAGCCT
3101 CCACTATCTC CCTGTGACGG GTGAACTTCG TGTACTGTGT CTCGGGTCCA
```

<p style="text-align: center;">FIGURE 1, page 1 of 2</p>

```
3151 TATATGAATT GTGAGCAGGG TTCATCTATT TTAAACACAG ATGTTTACAA
3201 AATAAAGATT ATTTCAAACC ACCAAAAAAA AAAAAAAAAA AAAAAAAAAA
3251 AAAAAAA  (SEQ ID NO:1)
```

**FEATURES:**
5'UTR:       1-44
Start Codon: 45
Stop Codon:  681
3'UTR:       684


**Homologous proteins:**
Top 10 BLAST Hits

| | Score | E |
|---|---|---|
| CRA\|335001101587561 /WO200058473 /org=Homo sapiens /taxon=9... | 428 | e-119 |
| CRA\|18000004937398 /altid=gi\|464561 /def=sp\|P35289\|RB15_RAT RAS... | 423 | e-117 |
| CRA\|18000005187045 /altid=gi\|7498104 /def=pir\|\|T33855 hypotheti... | 220 | 6e-56 |
| CRA\|18000004929618 /altid=gi\|131798 /def=sp\|P24407\|RAB8_HUMAN R... | 216 | 1e-54 |
| CRA\|335001098683352 /altid=gi\|11422744 /def=ref\|XP_001482.1\| TR... | 214 | 4e-54 |
| CRA\|18000005096141 /altid=gi\|2317272 /def=dbj\|BAA21744.1\| (AB00... | 214 | 4e-54 |
| CRA\|18000004952869 /altid=gi\|131848 /def=sp\|P22128\|RAB8_DISOM R... | 212 | 3e-53 |
| CRA\|335001098688905 /altid=gi\|11432830 /def=ref\|XP_007682.1\| RA... | 211 | 4e-53 |
| CRA\|18000004945380 /altid=gi\|131847 /def=sp\|P22127\|RAO1_DISOM R... | 211 | 5e-53 |
| CRA\|18000005163099 /altid=gi\|7705849 /def=ref\|NP_057215.1\| ras-... | 210 | 6e-53 |


BLAST dbEST hits:

| | Score | E |
|---|---|---|
| gi\|12333507 /dataset=dbest /taxon=96... | 626 | e-177 |
| gi\|12120217 /dataset=dbest /taxon=96... | 377 | e-102 |


**EXPRESSION INFORMATION FOR MODULATORY USE:**
library source:
From BLAST dbEST hits:
gi|12333507 brain
gi|12120217 epid_tumor

From tissue screening panels:
Fetal whole brain

FIGURE 1, page 2 of 2

```
  1 MAKQYDVLFR LLLIGDSGVG KTCLLCRFTD NEFHSSHIST IGVDFKMKTI
 51 EVDGIKVRIQ IWDTAGQERY QTITKQYYRR AQGIFLVYDI SSERSYQHIM
101 KWVSDVDEYA PEGVQKILIG NKADEEQKRQ VGREQGQQLA KEYGMDFYET
151 SACTNLNIKE SFTRLTELVL QAHRKELEGL RMRASNELAL AELEEEEGKP
201 EGPANSSKTC WC  (SEQ ID NO:2)
```

**FEATURES:**
**Functional domains and key regions:**
[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site

       205-208 NSSK

[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 2
    1      92-94 SER
    2   206-208 SSK

[3] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

Number of matches: 2
    1    29-32 TDNE
    2  104-107 SDVD

[4] PDOC00007 PS00007 TYR_PHOSPHO_SITE
Tyrosine kinase phosphorylation site

       101-109 KWVSDVDEY

[5] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 2
    1    18-23 GVGKTC
    2  136-141 GQQLAK

[6] PDOC00017 PS00017 ATP_GTP_A
ATP/GTP-binding site motif A (P-loop)

       15-22 GDSGVGKT

[7] PDOC00579 PS00675 SIGMA54_INTERACT_1
Sigma-54 interaction domain ATP-binding region A signature

       11-24 LLLIGDSGVGKTCL

**BLAST Alignment to Top Hit:**
>CRA|18000004937398 /altid=gi|464561 /def=sp|P35289|RB15_RAT
       RAS-RELATED PROTEIN RAB-15 /dataset=nraa /length=212
       Length = 212

```
 Score =  423 bits (1077), Expect = e-117
 Identities = 207/212 (97%), Positives = 209/212 (97%)
 Frame = +3


Query: 45   MAKQYDVLFRLLLIGDSGVGKTCLLCRFTDNEFHSSHISTIGVDFKMKTIEVDGIKVRIQ 224
            MAKQYDVLFRLLLIGDSGVGKTCLLCRFTDNEFHSSHISTIGVDFKMKTIEVDGIKVRIQ
Sbjct: 1    MAKQYDVLFRLLLIGDSGVGKTCLLCRFTDNEFHSSHISTIGVDFKMKTIEVDGIKVRIQ 60


Query: 225  IWDTAGQERYQTITKQYYRRAQGIFLVYDISSERSYQHIMKWVSDVDEYAPEGVQKILIG 404
            IWDTAGQERYQTITKQYYRRAQGIFLVYDISSERSYQHIMKWVSDVDEYAPEGVQKILIG
Sbjct: 61   IWDTAGQERYQTITKQYYRRAQGIFLVYDISSERSYQHIMKWVSDVDEYAPEGVQKILIG 120


Query: 405  NKADEEQKRQVGREQGQQLAKEYGMDFYETSACTNLNIKESFTRLTELVLQAHRKELEGL 584
            NKADEEQKRQVGREQGQQLAKEYGMDFYETSACTNLNIKESFTRLTELVLQAHRKEL+GL
Sbjct: 121  NKADEEQKRQVGREQGQQLAKEYGMDFYETSACTNLNIKESFTRLTELVLQAHRKELDGL 180


Query: 585  RMRASNELALAELEEEEGKPEGPANSSKTCWC 680
            R  ASNELALAELEE+EGK EGPANSSKTCWC
Sbjct: 181  RTCASNELALAELEEDEGKTEGPANSSKTCWC 212    (SEQ ID NO:4)
```

>CRA|335001101587561 /dataset=GENESEQ /org=Homo sapiens /taxon=9606
       /mol_type=protein /date=08-FEB-01 /length=218
       /altid=derwent_id|B41604 /altid=derwent_ac|B41604
       /def=Human ORFX ORF1368 polypeptide sequence SEQ ID
       NO:2736 /patent=WO200058473-A2 /pat_section=Claim
       Length = 218

```
 Score =  428 bits (1088), Expect = e-119
 Identities = 212/218 (97%), Positives = 212/218 (97%), Gaps = 6/218 (2%)
 Frame = +3

Query: 45   MAKQYDVLFRLLLIGDSGVGKTCLLCRFTDNEFHSSHISTIGVDFKMKTIEVDGIKVRIQ 224
            MAKQYDVLFRLLLIGDSGVGKTCLLCRFTDNEFHSSHISTIGVDFKMKTIEVDGIKVRIQ
Sbjct: 1    MAKQYDVLFRLLLIGDSGVGKTCLLCRFTDNEFHSSHISTIGVDFKMKTIEVDGIKVRIQ 60


Query: 225  IWDTAGQERYQTITKQYYRRAQGIFLVYDISSERSYQHIMKWVSDVDEYAPEGVQKILIG 404
            IWDTAGQERYQTITKQYYRRAQGIFLVYDISSERSYQHIMKWVSDVDEYAPEGVQKILIG
Sbjct: 61   IWDTAGQERYQTITKQYYRRAQGIFLVYDISSERSYQHIMKWVSDVDEYAPEGVQKILIG 120


Query: 405  NKADEEQKRQVGREQGQ------QLAKEYGMDFYETSACTNLNIKESFTRLTELVLQAHR 566
            NKADEEQKRQVGREQGQ      QLAKEYGMDFYETSACTNLNIKESFTRLTELVLQAHR
Sbjct: 121  NKADEEQKRQVGREQGQQKCPSLQLAKEYGMDFYETSACTNLNIKESFTRLTELVLQAHR 180


Query: 567  KELEGLRMRASNELALAELEEEEGKPEGPANSSKTCWC 680
            KELEGLRMRASNELALAELEEEEGKPEGPANSSKTCWC
Sbjct: 181  KELEGLRMRASNELALAELEEEEGKPEGPANSSKTCWC 218    (SEQ ID NO:5)
```

<div align="center">

## FIGURE 2, page 2 of 3

</div>

**Hmmer search results (Pfam):**

| Model | Description | Score | E-value | N |
|---|---|---|---|---|
| PF00071 | Ras family | 323.8 | 8.2e-95 | 1 |
| CE00060 | CE00060 rab_ras_like | 211.0 | 1.8e-59 | 1 |
| PF00006 | ATP synthase alpha/beta family | 4.2 | 2.1 | 1 |

Parsed for domains:

| Model | Domain | seq-f | seq-t | | hmm-f | hmm-t | | score | E-value |
|---|---|---|---|---|---|---|---|---|---|
| PF00006 | 1/1 | 10 | 24 | .. | 203 | 217 | .. | 4.2 | 2.1 |
| CE00060 | 1/1 | 2 | 165 | .. | 16 | 184 | .. | 211.0 | 1.8e-59 |
| PF00071 | 1/1 | 10 | 212 | .] | 1 | 198 | [] | 323.8 | 8.2e-95 |

FIGURE 2, page 3 of 3

```
   1 GCTCAAGATT GCACAGCTGG TGAGTGGTGA CACTGGGACT GGAACCCAAG
  51 TGTGCCTTAC TCCAGAGCCC TTGGCATGCA CCTGAAACCC CATGTAAGCC
 101 CACTGTGGAG ACGCGCACCT CGAAATAATG GAATCCACTA CATCAGTTCC
 151 TTTAGCTTTC TGTGTAATCA GAGTAGCTAG CAGGCTCGGG ATTTCGCCCC
 201 CCGGCTTTTT TTTTTTTTTT TTTTTGAGAC AGAGTTTTGC TCTTGTTGCC
 251 CAGGCTGGAG TGCAATGGCG CAATCTCGGC TCACCGCAAC CTTCGCCTCT
 301 CAGGTTCAAG CAATTCTCCT GCCTCAGCCT CCCGAGTAGC TGGGATTACA
 351 GGCACCGGCC ACCACGCCCA GCTAATTTTT TTATATTTTT AGTAGAGATG
 401 GGGTTTCACC ATGTTGGCCA GGCTGGTCTT GAACTTTTCC CCTCTTATTA
 451 TAATTCAGAC ACTTAACCTG AAATATACCT TTTCAAATGA AGTAAATGGG
 501 CTTACCACTT TCCTTGACCT ACTATTGAAA AATACATTCT CCATCCAATA
 551 TTCAGCCTGA AAACAGGTAT GTACATATAT ACTTTTCATT GCTTTTTTTT
 601 TTTTTTTTTT GAGACAAGGT CTCCCTCTGT TGCGCAGGCT GGAGTGCAGT
 651 GTCATGATCT CGGCTCACTG CAGCCTTCCC CTAATGGGTT CAAGCAATCC
 701 TCCCACCTCA GCCTCTCAAG CCTGGGATTA CAGGCGAGCC ACCGTGCCCA
 751 GCTAATTTTT TTTTATTTTT AGTAGAGACT GGGTTTCACT ACATTGGCCA
 801 GGCTGGTCTC CAGCTCCTGA CCTCAAAGTG ATCTGCCCGC CTCAGCCTCC
 851 CAAAGTACTG GGATTACAGG CATGAGCCAA CGCGCCTAGC CTTTCATTGC
 901 TTTTTAAAGA CCTAATAGGC TAGACTTTGC TCTCCCTCAA TACTCGTTGG
 951 TAGGGATAGG CAATTTTCTC AACTCCGGAG AGCATTCATT TGCCTCTCTC
1001 CGGTGCTAAC ACATTCAGTG GTAGGAAACT GGATCTTGAA CAAGGGCCAT
1051 TCATTCTTTG GTGCCACTGG CTATACCACA GAGAAATTTA GGGGTCTGAA
1101 ACAATACATT GGTCACCTGG GCACCTATCC TAAGCACCTT AGAGGGAAAA
1151 CGGAGACTTG CCCGCACACC TCTAAAGGAT TTTGCACTTG GAGATGTTCT
1201 TATGGCCATC TATCTTTTCA CCCTGGTGGA GGCCGTGAAT AGGCATTTTC
1251 CCCATTTAAA GAAAAAATGG GGACGGGGGA GGGCCGTGAC ACAGTCACAC
1301 AGGTAAGGGG CAGCCAGATG GCAGGGAGGG GGAATTCCAC CCACACTCTC
1351 GGGGACTCAT GGAGACGGGT GTTCGAATCC AGATCCTGCT CAAGGCCTTC
1401 CTACCTCGGG TGAGCCCAGC TGAGGTACCA GCCACTGGGG AGCCCGGCCA
1451 GATCCTGCAG ATGCAGGGTG CCACGGCGGG CGGAATTACC GGCGCCAGAC
1501 TTGGGGTGGG ATATGGGGAG AAGTGGTGAG CCCGGAAAGC GGAGCACGGT
1551 AGAAGTGGGC TGGGTGGGGG CTCACCTCAA CTCCCCCATT CGGAGCGTCC
1601 GCGGAAAAAC GAAAACGTTC CCCCGCCCCG GCAGGAAGG GGTTGGGAGG
1651 GGGGGCTGGC GCCCCGCCCC AGCGTCGCCT GCTCGATGGG GTCCCGCTCT
1701 CCTGCGCGCG CTCCCCGCCC CCTCTCTACC GGGGCGGCGG CGGCGGCGCA
1751 GGGGAAGGGG CGGGCAGGGG CCGCCGCCGG TTTCTCCTCC CACCGCCTCG
1801 CGCCAGCCCA GCCGAGCCGA GCCGAGCCGA GCGGGCGCCG CGCCGGGCTC
1851 CCGCCGCAGC CGCGCTTCCC GGCACCCAGC GAGCGAGTGG GCAGGCGGGC
1901 GGGCGAGGCA GCCGCGGGGG CCGGGCCCGG CGTCCTCCTC GCCGCCCGCA
1951 GCGTCCCCGG GCGGGCGCGG GCCGCGATGG CAGCGGCGGA GCAGGGCTGA
2001 GCCCGCTGCC CGCCCGCAGT TCCCGGCCCC GCTGGCCCCA GTCATGGCGA
2051 AGCAGTACGA TGTGCTGTTC CGGCTGCTGC TGATCGGGGA CTCCGGGGTG
2101 GGCAAGACCT GCCTGCTGTG CCGCTTCACC GACAACGAGT TCCACTCCTC
2151 GCACATCTCC ACCATCGGTA AGGGGCGGTG GCCCGGGGCG CCCCTCCCTC
2201 CCCGCCCGCG GCCCCTTTCC CCGCCGCCCC CGTCCCCAGC TGGGGAGGAA
2251 TTGCCAGCCC CTCCGGCTGG AGGCGGTGGC GCCGGAGGCC GGAGTCCGGG
2301 ATAAATCTCG GGGTGAGCAT AGGTTTTGGC AGGTGAGGGT GTCCCTGCTG
2351 CCTGCCGTCC GGACCAGGGG TGGGGTCTCC CGCCTCTTGC CGGGAAGCCT
2401 TCCGTCCCAT CAAACCGAGA AACCGGGGGT GAGGGGAGCT GGTGTAGGCC
2451 TGGGTACCCC GAGCTGGGGT AGCAAGAATC GTAGCCGCTG GAATAACACC
2501 CCCACACCCC CAGGGGGAGG GGAAGTAAAG CTTCTGCTAC GGAAAAGGGG
2551 GTCAGGGTGG AGACCGGAGT CACTGAGGCG CCCTTGGTTC TGTGGTGACC
2601 CAAGGTGGAG CCGGCGGGGG GCGAGGGGGG GAAGAGAGGA CGTACGGAGG
2651 GGCCACAGGG ATCGAGTTTC CAGGGCAGAG TTGGGAAGGT AAGCCGCAAG
2701 GTGGGACACC TGGGGGAGGA CACAGATAGG GTGAGGAGCC CCTGCGCCTG
2751 GGAAGAGGAG ACATCTGTTC TGAGGGAGGC TAAAGAGGAT GGAGGAGTGT
2801 CAGGAATACC TGCCCAGACC AAGGGGTCAG AAGGCAGGCA GGACCCGCCT
2851 GAGGGCATCT CTCATCTGGC AGTGCTGGAG CCTGTGGTTA GAGGGACAAG
2901 ACCCGGTGGC ATCCCAGACA GCACTATGAT GGGGTCACTT ATTCTAGGAA
2951 TGGGTCCATG GCCTCCCCTC TGAGACAGTC AGTCTCCCGC TTCTAGGCTG
3001 TGAGGGGCCC CCTCCCTGAG AAGTCTGAGT AGAGGGAATT TCATCCTCAG
3051 CTGCTACCCG GGTCAGCCCT GGAGTAGCCT CTGCATTGCC CAAGCCCCTG
3101 GAAACACCTG CTGGCTGGCT GGTCATCCAT TTGGAATGCT CTCCTAGAAG
```

## FIGURE 3, page 1 of 14

```
3151 TCCCTGCTGC CATCAGGGAT GGGCACCAGC TCTCAGCTTC CTCTTGAGGA
3201 TTCATGTCCA CACCATCCCC CCTCCCCCCA ACACACATTC CTTGCTGAGA
3251 GAGAAGTAGG AGCAGATAGA TACAGCCAGG AGGAACAGAA CCTTCTGGTT
3301 AAGAAGCCAG CTTTATTGTC CAAGAGACCT GAGACCTCAC TGTGGGGCAA
3351 AGCAACCTTG AATATTGCCT AAACTTCTGA GCTTTATTTA GTTTCTCATC
3401 TGTAGAACGG GTATAATAAT TGCACCTACC TGCCAAGTTG TTGTCAAGAT
3451 TAAATGAGAT AACGATTGTT AAGTGCTTAG CACAGCCAGA CACATGGTGA
3501 AGCTCGATAA ATGCTGATTG TTCTTACTGC TATTGCCATT ATCATTGAGC
3551 TTTTAGGGTC TCCTCTCTTT GTTTCACCAA CTTGAAGGGT GAAACAACAG
3601 GACTTAGGGT CAGGGAACAG AACTTGTCCG TCTTTCTCAG AGGAGCTGTA
3651 AGGCCAACTC TTAGGAAACC CAGGAGCTTG GGCTGAGCCA TGGTTTGGAT
3701 GAGAGACATT GCAGAAAGAA GGGGAGCCTA TAGACACTAA GGCTTTGTGC
3751 CTGCCGGGAG GACTTGGGGA AGAGGCAGGT GCAGGAGAAA GGCATGGGCG
3801 TGATGGAGGA AGTGGCAGAG GAACCAGATG GTGTATGAGG ACAGGTTGTG
3851 GGCTCAGGGA CAAAGGGCGG TGGGTTATCC CTTAAGGAAA CTAGGAGTGG
3901 CTATTTTTGG GAGAGGCCTG GTGCTTGGAA CTACTGAGCT ATCTCCAGAG
3951 AGCTGTGGGC TGCCTGGGAG GCCCTGGCTT TGGCCTGAAG AGCTGTTGTT
4001 TGCACCTGCT CTCCTAGTCC CATTCCAAGT CCTATAGGTG ACATGGACTT
4051 TTCCCTTTGA GGGCTTCATT CAACCACCTC ATTTCAGAAG CTCTGGGACT
4101 CCTGCTTAGT GCTGTGGGAG GCAGCCTCCC CTGGGAGACA CATACCCTCC
4151 TTTTTGAGGG CACCCCTCTT TCTAAAATAC CAGGATGGCC CTCTGAGGCT
4201 CGTGCTCTCC TTAAAGAGAG TCCATTGCCT CACACCTCTA ATCATCCACC
4251 CTTCTCCTTG TCCCTTCCCC TTGTAATCTC CCTTCTTAGA CACCTTCTGC
4301 TAATAGGTGA ACACTAAATA GGTCACAGGG ACTTCCTGAA ACCCTCCAGG
4351 GCAGACCACT TTGGGCACAT AGGTGAATCA GTGAACTGAG TAGGGGTGTC
4401 TCTGCAGCAC TGTCTCCCCT CAAGGCCCTT GGTATATTGG CCTAAAACCT
4451 AAAGATGGCT CCCAGATTTC TTCCTCCGCT TCTGACACCC GGGTTCCCCT
4501 TTCTACAGGA CACAGAGGAT TCTCTAGGGT CCCCCTTTCC ACAGGACACA
4551 GAGGACTCTA GGAGTTTGGA TTCCATGGAA TAGAAAGAAA CCTGTCTTTC
4601 TTCACACCAG CCTTTTAAAA TCTGCCCCAC TGGGTATCTT AAATGCTTTC
4651 TTATTTAAAG CTTATTAAGG GACTTGGGAT TCTCCCTTAT CTTGGGCGTG
4701 TTTTTCAGCA TTAACTAAAA CTTAAAGGAA AGAGTTGGAT GGTCAAGAAA
4751 AGCTTTTTCC TTAAGTGATA TGGACAGTTT CTCAAGGAGG TAGAAGGGGC
4801 AGCCAGGAGA CAAATCAAGG AGCCAACGAA ATGAGTGCTA CCAAGTCATA
4851 GTCATTCGCT TATTTTTAAA AAATGCGTGT CCTGTATGCC AGGCTCTGCA
4901 CTGAGACCGA GAGATTCCAA GATGAATAAT ACCTACAGTC ACTGTTCTCA
4951 AATTGTGCAT TACCTAAAAC ACATTACATG ACCATGCTGG CCACTGATCG
5001 AGGCACCTTT CCCAGGGGCT TTTTTTGTGA ATTAAGAAAA CAAGGTAATT
5051 CACCAGTTAT TGCCAAGATA GTTTGGCTTC TTGGCTCATG TGGATATCAC
5101 CTAGGCCAGT ACTTTTGTGA TTTACTGTGT ACTCCACTTT AACGGCCTGC
5151 GATCTTCTAG AGAAGAACCC GCCAGGGAGC AGTGAGAGGC CTCCCTGGTA
5201 GACTGAGACA CTGACTGTCC CTCCCCCTAT CCTTTTCGTC TTTCTGGCCA
5251 GCAGACCAGC AGGTGGCCCT GCCACTGGCT CTGCCACAGG CATTTCCTTT
5301 CTGTGCAGCT GTGCTGGCCT GGCTGGGGGT TGGTGCGAAG GGGTCCCCAA
5351 AATACTACCT TAAACAAATT AATTGAGCAT TCACTACCAA GCTCTGTGCC
5401 AGGCATTTTA GAGACATATT GCAGTCTACG TTTTCTGCCC ACAGAAGCCC
5451 ATAACCTAGA TGGGGAGGCA AGACAAAGGG AAAAACAAAA AACAAAGAGC
5501 TAGTGCCAAA ATGAGATATC TGAAAGAACT TGGTGAATCA CTCTTCAAAT
5551 GTAAAGGATG GATTATGATC ATTGCAGTTA CTCTTAATGA AGGTCTCACA
5601 GTGGGTATCA GAAGCTAAAT TATGATGCAA GATGTACCAT GAGGCAGCCG
5651 GAGAATGGCG ATGGATGGGA TGGGTGAGTG CTATTCCCAC GACTCCATGC
5701 TGTCGGAGGC TGGGGAAGAG AGAGGCCCCT GTGGACTAGA ACCGGCAGGG
5751 AAGGCTGAAG CTAGGCCTCA GTGTGGGCTG CTCGTCAGTT CCTGCAGCAG
5801 AAGGGAGCAG ATGGAGTAAC ATGAGCAGAG ATAACAGAGG TGGGATTGAG
5851 TAGGTGTCCG TGGGGCTCTA GGCAGTTTAG ATGCAACAGA AGGGATTCTT
5901 CAGGAAAGTG AGAAGATTCT TCTGTTTCTC TCTCTGTCTC CCAAATTATA
5951 AGTGCCTTGA TGGTGCGACC AAATCTTATT CCTCATTGTT TTTATAGTCC
6001 CTAGTACAGG GCCAGGCAGA TTCAATGCCT GTTGTTAAAT TAATGAATGA
6051 ATGCAGGGAC CAGTTGGCAG AGGGCATTGA GAGCCTGGCC AAGGAGGTGG
6101 AACATGAGCC TTAGCAATGG TAGGAGGGGT TTTGAGTAGG GTACTAATGA
6151 GGTTGGCTGG AAGAAGGGGT TAAGACTTGA AGCAGGGAGA CTAGTCAGGG
6201 GCTGCAGTAG TATCCTGGGC ATGAAGGAAC CTCTGAATGG CCCCTCACCC
6251 CCAGTGGTAC CAACACCAAC TTCCACACAG TCAGTTGTTC TACTTTCCCT
```

```
6301 CCAGAATGGG GAGTGGTTCA AGCCAATCAA CCTGGCAACT TCTGAAAGAA
6351 TCTTATGGGA CCTGTGCCAT GACCAGGTAG GGAGAAGATG TCATACATGG
6401 ACATCTATGT TCAGGGGACC TTTGAGGACC TTTCTGCATG GTGGCCAGGA
6451 GTGAGATGAT GTAAACCACA AATGGAAACT GAAGAGACTG CTCAGGAGTT
6501 GTTGGTTTTC TTTTCTTTTC TATTTTTTTT TTTTTGAGAC TAGGTTTCAC
6551 TCTGTCACCC AGTCTGGAGT GTGGTGGTGG CACAATCACG GCTCACTGCA
6601 GCCTCGATCT CCTAAACGCA ATCCTCCCAC CTCAGCCTCT CAAGTAGCTG
6651 GGACTACAGG TGCATGCCAC CACATTCAGC TAATGTTTGT ACGTTTTGTA
6701 GAGATGGGGT TTCACTATGT TGACCAGGCT GGTCTCGAAC TCCTGGACTC
6751 GTGATCCACC AGCCTCAGCC TTCCAAAATG CTGGGATTAT AGGCGTGAGC
6801 TACCTCACTC CCTCAGGAGT TGGTTTTCTC CCTCCCATCC TTAGTCTTCC
6851 CTGAGTAGAC CTGTCACCTA GTCCCTGGAC CTTTTGTTTT GAAAGCCACC
6901 CTCCAGGCTA CACTCCTTCT GGGTGAGGAG GAGGGTGATC TGGTTGGACA
6951 GGTTGGGCTG CTGTGGCTTC AGGGCACTTT CTCAGGCTGG GTTGCTGCTG
7001 CTATGTCACC TTTCTCAAGG AGTTCTGCTG GGACTGGCTT GGCTGCCTGT
7051 CTTGACTTTG CTTTTGACTG AGGAGGTGGG AGATGGTGAG GGAGGGGGTG
7101 GGGCTAGATC CAAGCCTGGA ATGGGGTGAC CTAACAGACA CTGGGGCCTG
7151 TGCTTAGACA CTAGGATCCT GGGGTTTGCA GGTTTCTAGA CTGAGAGGAG
7201 CTGGGGGCAA ATGCAGTGTG ACGTTGTGAG AGGGTCAGGG CTGGGTCTGT
7251 GTCAGCCTTC AGGCAGCCTG AGACCAGTCT CTACCTACTC TGTTCCCCTG
7301 GTACCTAGAA AGGAAGGGAA GGTGAGAAGC AATGAGCAGA ATGGAAAGAG
7351 CCCAGATTAA CATGCACATT TCCCATGGCC TTACTGGCCC TGTGACCTTC
7401 AGACACTTTG ATGACATCTT TGTGCTTCGT TTCTGCATCT GTAAATTGAA
7451 GATGGTAACA GAGTCTTTCT TAAAGGTTGT TGTGAAGATT ATAGAGCCTA
7501 GCGCATATAA AGCACTTGGC AGAGCCCTCG ATAAAATAAT AGCTGCTATC
7551 ATATTATCAT TATTATTATT TTATTTATTT ATTTATTTAT TTTTTTTTGA
7601 GACCGAGTAT CTCTCTGTCG CCCAGGCTGG AGTGCAGTGG CACAATCTCG
7651 GCTCACTGCA ACCTCCATCT CCCGGGTTTA AGTGATTCTC CTGCCTCAGC
7701 CTCCTAAGTA GCTGGGATTA CAGGCACCCA CCACCACACC CGGCTATTAT
7751 TATTATTCCT AGCTATAAGA ATGCTGTAGA GATGAATACA CTGTCAGTGA
7801 GCTAGGAGGT CATCCTGTGT ATCCATCACT TGTGCACTCA GTCGTTCAGG
7851 CGCTATTTGC TGAACACCAA CTACATGCCA GGTGCCATGC TAAGATTTGG
7901 GGACACAGTG GTGACCAAAA CAGACAGAAA CCAAGGAGCT GGCTTACATT
7951 CCAAGGGAGT GCATAGGAAG CTGTGTTTCA TTTCAGTTTC TGCTCTAGTA
8001 CCCCCCTTTC CCTGGCAGTG CCAGGGTCTG AGAAGGAAGA GTGAGGTGGT
8051 GAGGAGGTGT GAAGCAGTGG GGTGACCTGA GAGGAGAGGA TGGGGTGGCT
8101 TTGCCTCAAG GCTTGGGCCC CTGCTAGGTG TCGCTCTGCC TCAGGCCTCT
8151 GTTTCTCCTC CTGACACAGG CACAGACTCG GCCTCCCACC CCTTCCCCAA
8201 GGACATGACC TTGGGAAGGA ACATATCTGA AGCCCGCGGA GGGTTTCCGC
8251 TGCTGTGCAT CTGTGCCACA GATCCGCAGA TGCACCCACA GCTGGGAGCA
8301 CCGGTTCCTC CCGCCTACCT GCACTCCCTG GTTTCTGTTC CTTCCTCCTC
8351 CTCCTTCCTT CTCCCCGCTC CCCAGACAGG CTGGTGATGA GCTTTATAAC
8401 ATGAAAGCTG ATATTTGGCC ATTATCCTTC TACCCTGATT GCCAGCTCTT
8451 CTCAGAGTGC CTTCTTCTGT AATCCAATCT TTGCACCAGT TTCCCTGTGA
8501 AACTGCCAGT TTTCTGTATA GGCCTCTGCC CTCTCCTTGG CTCTTCTCTC
8551 TGGTCAGTGA GCTTTGTCAA GGGGAACACA GGGCTTCCTG GACACGTAAT
8601 TCCTCCCACT GAGGAGGAAG GGGCTAATCA CCAGCCCTGT TTTATTTTAT
8651 TTTATTTTTT TGAGATGAAG TCTAGCTCTG TCGCCCAGGC TGGAGTGCAA
8701 ATGGCTCGAT CTCGGCTCAC TGCAACTTCT GTCTCCCGGG TTCAAGCGAT
8751 TCTTCTGCCT CAGCCTCCTG AGTAGCTGGG GATTACAAGC ATGCACCACC
8801 ACACCTGGCT AATTTTTTGT GTTTTTAGTA GAGATGGGGT TTCACCATGT
8851 TGGCCAGGCT GGTCTCGAAC TTCTGACCTC AGCTGATCCA CCCACCTCGG
8901 CCTCCCAAAG TGCTGGGATT ACAGGAGTGA GCCACCATGG CTGGCCGACC
8951 CCATCTCTTA AAAAAACAAA AAGAAAAGAA AAGAAAACAA AACAAAAACA
9001 CTTTTTAAAT TAACTGATTA TGGTGGCATG TGCCTGTAGT CCTAACTACT
9051 CAGGAGGCTG AAGTGGAAGG ATTGCTTGAG CCCAAGTAGT TGGAGGCCAC
9101 AGTGAGCTGT GATCACACCA CTGTACTCCA GCCTGGGTGA CAGAGTGAGA
9151 CCCTGTCTCA GGAAAAAAAA AAAATTACTG AGAACTCTGT GACCATGGCA
9201 CCATGAACTA TAGAAAGGGC TAACAGTTGG CTTTGAAATG TGGGTTATGG
9251 CTGGGTGCGG TGGCTCACGC CTGTAATCCC AGCACTTTGG GAGGCCAAGG
9301 TGGGCAGATC ACAAGGTCAG GAGTTTGAGA CCAGCCCGGC CAACATAGTG
9351 AAACCTCATC TCTACTAAAA ATACAAAAAA TTAGCCGGGT GTTGTGGCAG
9401 GTGCCTGTAA TCCTAGCTAC TCGGGAGGCT GAGGCAGGAG AATTGCTTGA
```

```
 9451 ACCCAGGAGG TGGAGGTTGC CACAAGCTGA GATCGCACCA CTGCACTCCA
 9501 GCCTGGGCGA CAGAGCAAGA CTCCATCTCA AAAACAAAAA TAAAAACAAA
 9551 AAAAAGTGGT TTGTTTTCTT TTCTTTCTTT TTTCTTTTTT TTTTTTTTTT
 9601 TTTTGAAACA GAGTCTTGCT CTGTCACCAG GCTGGATTGC AGTGGAGGAT
 9651 CTCAGCACAC TGCCACCTCT GCCTCCCAGG TTCAAGTGAT TTCCCTGCCT
 9701 CAGCCTCCAG AGTAGCTGGG ACTACAGGCA CGCACCACCA CGCTGGGCTA
 9751 AGTTTTTGTA TTTTAGTACA GAAGGGGTTT CACCATGTTG GCCAGGATGG
 9801 TCTCCATCTC CCTGACCTCG TGATCCGCCC ACCTCGGCCT CCCAAAGTGC
 9851 TGGGATTACG GGCATGAGCC ACCACGCCCG GCCTAAAAGT GGGTTATTTT
 9901 CTAATTGCTC TTCCCTGATT AAAATTTTCT CTTTGCCCAT CTTTTCTCTA
 9951 GATATGTACT GACTTCATTC ATCCATTTAT TCGTCTCACT TGCTCATTCA
10001 TTTTTGCTTT CATTTACTCT ACTTTGTTGA ATAATATTTA GTGATCTACC
10051 TGCTGCCAGG CAGTGAGAGT CTGAAGTGAA CAGGATGCTG CTTTGCCCTC
10101 TGGGAGCTTA CAGTGTAGCT GGGAACCAGA CATCCAAACA AGCAGAATAT
10151 TATGCAAAAG AAATGTCAGG ATGCTTTGGA ATCACAGAGG AGTGAGAAAT
10201 CCCTCCCGGG GAGGCTGGTG AAGGCTTTGA AGAGGAAGTG ACATTTGAGT
10251 GGAGTCTTGA AGACTAGGCA GGATTCTCCA GGGGCCCTGG GTGTGGGGGA
10301 AGCACACATC CTCTTCCCTG TAGGAGGTGC TGTGGAGAAC ACCTCCAGTG
10351 GGGCTGCTAC TCTTCAGCCT TGCTGGGGCC AGCTGGAGTG GCCACACCAT
10401 GGTCACACCA GCTGAAGTTC AAGAAGCCCC TTGCCAGGAG ATTGCTTTGC
10451 TGGCTCTGGG TGAGGGCAGG TGCATCTGGA AGCCCCCTTC TTTCTAAGAT
10501 GTTTGCTCCT GAGTTTCTAT GTCCTAGTCT TTTCTTCCCT GAACCTTTTG
10551 CTACCAGTCA GCACAGCCCT GCCTGAGAAG GAGGCTGGAG GAGTGAGTGG
10601 TCAGTAGCCT GGTGGGTCTT GGCTGCCTCT GTGGTGCCCG CTGGCCTAAG
10651 TAGCAGGCTT AGGGAGGCGA GACCCAGTTC CAGGGGCTGC CAATGGGGAG
10701 CGAGATGGGG TGGCTGGAGC ACACTGCACA TGTCACCAAG GCTCTAGGGA
10751 GGTCTGTGCA CAAGGCAGTG GGAAAAGCAA GGGGAAGACC CAGCCTGGTC
10801 AACATGGTGA AACCCGTCT CTACTAAAAA TACAAAAATT AGCTGGGTGT
10851 GGTAGAGCAC GCCTGTAGTC CCAGCTAACT TGGGAGCCTG AGGCAGGAGA
10901 ATCACTTTAA CACAGGAGGT GGAGGTTGCA GTGAGCCGAG ATCGTACCAC
10951 TGTACTCCAG CCTGGGTGAC AGAGTGAGAC CCTGTCTCAA AAAAAAAAAA
11001 AAAAAAAAAA AAAAGTGGG GAAGGGGAAC ACTGATCCTG ATTATCTACT
11051 CCATATACTT ACTATGTACC TACTACCTAC ACAGGGACGG TGGGCTTTAC
11101 GCATGCCATT TATTCAGTGT ATAGAGATCT CAGCATCACA TAGGAGCAGG
11151 GAGTTCTGAA GTTGGCCTTG CTGGCATTTG AGAAGTTTCT TGGTGTATTC
11201 TTCAGGTTCA CGCCTCCAGA CAAGTGTAAG TGCTATTGAA TGCTGACTAT
11251 GTTCCAGGAA CTAAACCAGA TGCTAGAAGA CACGCAGTAA ACAGTACAGA
11301 TGCAGGTGCA CATGTGAGGG TCCACACAAG ACCTGAGAGA AGGGAGGGGT
11351 CTTGCTGCAG TTCCCCTTTT GTAACAAAGG AGAGAGTACT GTTGACCCTC
11401 TTCCTAGGAA CTGTGAGTTT GACTGAAATG TGTCCTGCCA CAGGATCTTT
11451 GCTGCTTCCT CTACCTGATT CTTTGGATCT CCCTGCTGGC ACCTTCTTGT
11501 CATTTAGGTC TCAGCTCAAA TGTTACCTCC TTTAAAATGT CTTCTCTGGC
11551 CAGCCAGTCT AAGGTTGCTT GTGCTTGGGG TCTCCTCACT CTCTACTTTA
11601 TCCCGCAGTT GCTTCTTATC ACATATGGCT CTCTGAAATT AGGTATTCAT
11651 TACTTACATC TGTCTTCCCC ACTAGAATTA AGCTCTGATG ACAAGGATCT
11701 TTCTGTGCTG TTCATAGCTT ATCTTCTAGT ACCTGGCTTA GTTCCTGGCA
11751 CATTGTAAGC ATTCAATAAC AGTTTGAATG AATGAATTAA CAAATGAAGG
11801 AATGAATGAA TGCATTTTCC TAGAGGACTT CTGTTCTTCC CTGAGGGAAG
11851 TTATAGGTCG TATTGGTTTC TTGGGACTGT TTTTTGTTTG TTTGTTTTGT
11901 TTTGTTTTTT GAGACAGAGT CTCACTGTAT CCCCCAGGCT GGAGTGCAGT
11951 GGCACAATCT TGGCTCACTG CAACTTCCGC CTCCCAGGTT CAAGCGATTC
12001 TCATGCCTCA GCCTCCCGAG TAGCTGGGGA TTCCAGGAGC CTGCCACCAC
12051 GACCAGCTAA TTTTTGTATT TTTAGTAGAG ACAAGGTTTC ACCATGTTGG
12101 CCAGGCTGGT CTTGAACTCC TGACCTCAGG TGACCTGCCT GCCTCTGCCT
12151 CCCAAAGTGC TGGGATTACA GGCATGAGCC ACCACGCCCG GCCTGTTTTT
12201 TTTTTTTTTT TAAGACAGAG TCTTGCACTG TCTCCCAGAC TGGAGTGCAG
12251 TGGTGTGATC TCAGCTCATT GCAGCCTCAA CCTCCTGGCC TCAGGTCCAG
12301 GTGATCCTCT TACCTCAGTC TTCTGAGTAA CTGGGCCCAC TGGTATATAC
12351 CACCACACCT GGCTAATTTT TAAATTTTTT GCAGAGACAT GGTCTCACTA
12401 TGTTGCCCTG ACTGATCTTG AACTCCTTGG GTTCAAGTGA TCCTCACACC
12451 TTGGCTTCCC AAAGTGCTGG GTTTACAGGT GTGAGCCACC ATGCCTGGGC
12501 TTGAGACTGT TAAGATGATG AGGCTGGAGG GAGTGGATGG CCTCACTGCT
12551 TGAGCCCTAG AGATTCCTTA CTCCAGAGTG CCCTGGCTGC AGAGGTGGCC
```

```
12601 CTGGAGGGTC ACTCCAGCAA CCTGGCTGAG CTGATGGGCA TCATCTGATA
12651 CCAGCTCTGA CCCTGAATAA TAGGCAACAT GGACCTTAGT CTAAAGCACT
12701 GACCCCTCAT CTCTGCATAT ACCAAAGAAG ATGAGATTTG GGTGAGGACA
12751 CAGCCAAACC ATATCAGCTC CCGGGATCCC TGTGTGAATG GGGTCTTTTT
12801 TGGTGTTTGA GGGCTGCACA GGGTGACCTC TTTAGAGGTG ACCTCCTGCC
12851 ACAACCCACA GGAGGTGCAC ATGGCCCACA CATGCTGGTT TCCTGCAGTG
12901 GGAGGGGCTG GGGCACTCCT GGGACCTGTG CTTGGTAACT GGAGCTGGCC
12951 TGGCCCTGGG GATTGGGTGT CTGCCTTGGG TTTCAGGTGT ATTAGGTTGT
13001 TCCTCGTTGT GGAGTCTCAT TACTAATGAA AAGTTCAGGT CGCACTGCTG
13051 GTCCTTTGGG CTGTGGTTGA TCCTGGTGAT AACATTTGGC ACCCAGAGGC
13101 AGCCCTGTTT CCACTGAAGC ATGCGGAGCT TGGCTGGCAG GCAGGCAAGC
13151 TGGCAGCTGC CCTTAACCCA TGAGGTGCTG GCCCGCTAGT AGGCACACCC
13201 TACCTGTGCC AGAATTGAGG TTGTAGCCAG ACTCCAGGAG CCATCTGGGC
13251 CCCACAGGGG GCGGCATTTC CTCTTTTTGT TGAAACATTC CAGCCAAGTG
13301 CTGGCTTGGG CTTCATCTCT CTGTCCCACT CTCCTTCCTC TCCCCAACAT
13351 AAGCCTCCTT CTACATCCTA GAGCTCTTTC CATTCCCCCT CCTGCAGCTC
13401 TGGGCTCGCT AATCTCATGC TTCCCTAAGG GAGCTTGACG GCTGCTTCTG
13451 CTAACATTTA ATAAAGTTCT GCGTGCCAGA CCCTGTGTTA TGGGTTTTAC
13501 ACCTTATCTC ACAATCTTAA AAAAAAAATT CTCTGAGAAT CCTCTGTCAC
13551 CCCCACTTTA CAGGTGAGGA AACTGAGGCA AAGATAGGCT AACTGGCTTC
13601 CCCAACACCA TGCAGGTAAT TAGTGATAAA GGCAGGGTTG GAACCAAACT
13651 TGACCTCCCA ATTGTGCTCT TAATGGCCAG GACACTCTGT GTCTTGAGCC
13701 ACACTTCCTC CATGTTTTCT AGGGCTTTCT AGGGAGGCAG ACAGTGATGG
13751 GAAGGGGTGT TCTTTAGTGT GGATGTGCCC TGCCTGCTCC TTTCTGTAAG
13801 CGTCACAGCA CCTCCACTGC TGTACTGGGG AGGCACCAAG TTTTTCCCTG
13851 TTTGCCCACC CAAGGCGAGC TAGCTTAGGA GTCACGTGAG TGCTGGGTGT
13901 CTCGCCTGCT GCATCCCTCT ATCCTGCCCC TGCCCCCGGT GCCCAGAGGA
13951 GGGCCCTGCC TGTCTTCCCA GTTCTCCAAC AGCAGCGCTG TCCCAGCACC
14001 CTCGGGCTCC AGTTGTGGCC TGGCAGCTGC TGGGGCAGAC ACCATACAGA
14051 CAGAGTCACA GCAGGAAGAG GATGGGGCCC AGGGCTGCTG CCTCAGGCCA
14101 TGGCTGCATG GCACCATCAG TTGATTGAGG AGCTTTTCTT GCCAATGTCT
14151 GAGGCATCAG GTGGCAGGAC ACGTCTCCCT GCTCTTAAGC CTCAGGCATG
14201 CAGCCCTTCT TATGCTCTCT GGGGTGAGGG GGAGATCCCC CTCATGGAAT
14251 TGCTTTTTTT TTTTTTTTTT TTTTTTTGAG ACAGGGTCCT GCTCTGTCAC
14301 TCAGGCTGGA GTGCAGCCTC AACCTCCCAG ACTCAAGTGA TCCTCCTGCC
14351 TCAGCCTCCC GAGTAGCTGG GACCACAGGT GGACACCATC ACACCTGGGT
14401 TTTTTGTTT TTTGTTTTTT GTTTTCTAGA GATGGGGTCT CACTTTCTTG
14451 CTCAGTCTGG TCTCGAACTC CTGGGCGCAA GCAGTCCTCC CACCTCGTCT
14501 TCCCAAAGTG TTTGGATTAC AGGTGTGAGC CACTGTGCTT GGCCTTTTTA
14551 TTTATTTAGA ATTTGTTTTG GAATTGCTTC TTTATGCCTG GCACTATGCT
14601 GGCACTATGT GGCAGAGATT TTAAAAACGA GCAAACAAAA CAAATGCTTT
14651 GTCAACCACA AAATGTATTC TCTGCCCCTT AGGTTCTTTT TGTGTAGTTG
14701 AGGCTAGAAG ACAAAAATAG GGGGCAGTAA GGAGCAGGGA GCGATGGTTT
14751 AGGAGGTCTT CCTTCCAGCC CCCTTGTTGA AGCATCTGGC TCACTAGCTT
14801 GGGGGAGCCA TTAGGCAGCA GTGGCCAATC CTGAGGCACT CTCAGGTGTC
14851 ACTAAGAAAA GGGGCATGTG CTCTATGGAT ACCCATGGGC TGAACTTGGA
14901 GTCTGGTCTG GACCCATGGC TGTGCTAGGA TCCACCGTCC CCAGCCCCAA
14951 CTGCAGTCAG CATGTTCATC ATCCTTAGGC CTCTCCGCTT CTTTCTGCAT
15001 GTTTGTCTGC CTCATGCCCT GCTCATTACC AACTGGTCAG TCCCCACTGC
15051 CCTGCCTGGA GTGAGCTGGT TTGATTGGCT TAGCTAAGCT CCCTTGCCTC
15101 TGCTGGCCAG GTCACCCTGT GGGTCACCAG CAAACCTGTT GATGGTCCAG
15151 TCTGAACCTG CTTCTCCACA AAGAACATGT TGCACCCAGC CCTGCTTCTC
15201 TGAGCAGAGG TTTGGGGCTG AGCTGTTCTA GCCAGAAAGG GACACAGGGT
15251 GTGGCAGGCA CCATGATGGG CATATCTAAT GTGCCGGGAA AAACAATGAG
15301 CTGCTCTCCG TGCTTTGGGC ACCTGGTTGG GAGAGGGCCC ATCTGTCTGA
15351 CTTTCTCCTC CTGGGGCTCT CAGCGTCTCC GAGAACCTCT GCCAGAGCTG
15401 TGTAGAAGTG GTTTGCTTGT TTCTTAACAC TTCTGTGCCC TATTTCTTTC
15451 TGTACCCAAG AAAGGAAGTA GACTGTTTTG TAGGGACACT GTCGGGGTGA
15501 TGAATCTGGA CTTACTGGAA TCATGAACCA TGCCAAGGAG GAAGGAGAAA
15551 ATAGGCTATG GTGGGTGTCT TAGTTAGGGC TGGCTGCTGT AACAAAATGC
15601 CTTTAGCTGA GTAATTTAAA GCAAGAGAAA TGTATTGCTC AGAGTTTGGG
15651 AGGCTGGGAA GTCCAAGATC AGGGTGCCAG CAGATTCAGT GTCTGGTGAA
15701 GGCTGATGCT CTGTGACAAA GGTGGCACCT TCTAGCTCCA TCCTCACATG
```

```
15751 GCAGAAGAGG GAACAAGCTC CCTCAGACCT CTTTTCTAAG GGCGTTAGTC
15801 CCATGCATGA GGGCTCTAAC ATCACGACTG AGTCACCTCC CAAAGCCCTC
15851 ACCTCCCACC AGCACTGCAC TGGGGATTAA GTTTCAATAT GGGAATTTTG
15901 GAGGAACACA GACCTTCAGA CCACAGCAGC GGGCTTCTCC TCATGTGCCC
15951 CCTGCCTCAC TTCTAGATGC CGCATAATGT CAGTGAAACC CCGTCTCTAC
16001 TAAAAATACA AAAAATTAGC TGGGTGTGGT GGCACGTGCC TGTAATCCCA
16051 GCTACTTGGG AGGCTGAGGC AGGAGAATCG CTTGAACCCA GGAGGCAGAG
16101 GTTGCAGTGA CCTGAGATCG TGCCACTGCA CTCCAGCCTG GGCGACAGAG
16151 GAAGACTCCG TCAAGAAAAA AGAGAAAAGG CATCAGGTAT GCCAGGGTGT
16201 GCGGGAAAAG GCATCGGGTA TGCCAGGGCG TGTGGGAAAA GGCATCGGGT
16251 ATGCCAGGGT GTGTGGGAAA AGGCATCGGG TATGCCAGGG CATGTGGGAA
16301 AAGGTGGTAA GATTCCTCAG CCTCCCAGGG TTGGGAAGCC TCTGGCCGAG
16351 TGAAGCATAC CCTGGGTGGG TTTTAAGACA CCAGCTTTCC AGTCCAGCTC
16401 AGCTGTGGGA TGTGGGAACA TGAGTCAGTG GGAACATGAG AATTGGCTTC
16451 CCTGTGGCTC ACAATAATAC CTACTCCTGC CTACTTCATG GGACCCGCAT
16501 AAGAGCTGAG GGATTCCATA GCTCAGGGGT ATGCTGTAAA GACAAGCACT
16551 ATGCACCTGG GTGTGGTTCT GAAACTTTCT TGTGCAGAAG AGTGAGTAGG
16601 GCTGGGCGAG TCCTGAGAAT GTGCATTTCT CACACACCTC TGATGCTGCT
16651 GATGCTCTAG TCCCTTGGCT GGCAAGGGTA CCTGGTTAGT AGGGGCCAGG
16701 ACTCTGTAAT GCCTTCCACT TCAGGGTTCT CTGGGCTGGT TTTCCTGACT
16751 CCCCAGGAAG CCTTTATTCA GCAGAGGGAA GGTAGGAGTG AGAGGACTAC
16801 GCTGTCAGTG CTTCACATAC ATCGTTTAAT TTATCCCAGC ACAGCCCTTA
16851 GGAGGGAAGC AGTATTCTCC TTCTACACTT AAGAAAATCG GCCTGGTGCG
16901 GAGGCTCATG CCTATAATCC CAGCACTGTG GGAAGCTGAG GCGGGAGGAT
16951 CGCTGGAGCC CAGGAGTTCA AGACTAGTCT AGGCAATACA GGGAGACCTC
17001 ATCTCTACAA AAAAAAAAAA AATTAGCTGG GCATGGTGGT GCACACTTGC
17051 AGTCCCAGCT ACCTACCCAG AGGCTGAGCT GGGAGGATTG CTTGAGTCCT
17101 GGAGGATCGA GGCTGCAGTG AGCTATGATT GCTCCACTAC ACTCCATCCC
17151 TGGCAACAGA GTGAGACTCC ATCCCAAAAA AAAAAAAAAA TTGAAGCTAG
17201 GAGAAGTTGA GACTTGCCTG AAGTTACACA GTAAGTGCCA GAACCAGGAC
17251 TTGGACCAGG TCTTTCTGAC TCCAGGCCAA TGGATGTTTC TTCCATGACA
17301 TATATAGCTC TTGAAACTAC TTCTATCTAA TATCACCCAC AGTGCTGTTA
17351 AAAATACAGA TTTCTGGGCC TCACCCTCAA ATTATGATTC AGTAGGTCTA
17401 GGCACGTCAA GGTCATTGTT TTTGTCTTTG TTTTAAGTCA CCCCAGGTGA
17451 TTCTAAAGCC GAAGCTCTGC AAAGCACACC TTGAGAAACA GAGAACTCTT
17501 GTGCTCTCGC TCTCTTGACA CTTCAGGTGC AAAACTTTTG TCCTAATGTC
17551 GTTCTCAAAC TTACGCATGT GTGAGAATCA CTGTGAGAGC TTATTGAAAC
17601 TGATTGCGGG ACCCCATACC TAGAGGGCCT GATTCTATAG GTCTGAGGTA
17651 AGGCCCAAGA ATTTGCATAT TTGCATTTCG TTTTCTTTTC CTTTCTTTTC
17701 TTTTTTTTTT TTTTTGAGAT GAAGTCTCAC CCTGTCGCCC AGACTGGAGT
17751 GCAGTGGCAT GATCTCAGCT CACTGCAGCC TCTGCCTCCT GGGTTAAAGC
17801 GATTCTCCCC ACACCCCAGA CCCGCTCCTG AGTAGCTGGG ATTACAGGTG
17851 CCCGCCACCA TGACTAGCTA ACGTTTGTAT TTTTAGTAGA GACGGGGGTT
17901 TCACCATGTT GGCCAGGCTG GTCTCAAACT CCTGACCTCA GGTGATCCAC
17951 TCACCTCAGC CTCCCAAGGT CTTGGGATTA CTGGTGTGAG CCACCGCGTG
18001 CGGCCAGAAT TTGCATTTCT AACAAGTCCC AGGTGATGCT GATGCTGTGG
18051 GTCCAGGGAC ACACTTTGAG AACAGCTTGT TACTCAGGCG ATATGTGGAC
18101 AGTAGCGTCA TCTTCACCTG GGAGCTTCCT GCAGCATCTC AGGCCTTGCC
18151 CTACACCTAC CAGATCAGAA TCTGCATTTT AACTCAATCC CCGCGTGATT
18201 CTCATGCACC TGGAAGTTTG AGAAATATGA CCTTAGAGGA GCCGGAATGT
18251 GAAACCACTG GAGGCAGAGA TAGATGGAGA ATATCTCTTC TTCTCACGGA
18301 TACTAAAGAT GCAACAAAAA GGGCTGACTC TCTGGGTGTG CACCCAGGTG
18351 GGGCTGATGA CCGAAAAGAG GCCAGATGTG GACAGAGGAC TCTTCCCTGA
18401 GGGAAGGCAG AGAGAACTTA GGAAAATCTG AAGAAAGGAG GTGGCTTCAG
18451 AGGAAAGGCA TTCATCTGGG CCATAAAACA GTGGAGAAGG TATCCTGCTG
18501 AGAGCACAGG GGTGGGGAGG GGGTGCCCTG GAGCTGAAGT CTTCAGTGGG
18551 GGGACAGTGA TAGGTGAACA CACATGTGAA TAAACAGTTT GCTAAGCAGC
18601 TGCGAGGGCT GGCCAAGGTG AGAAAGCATC CGTCTGCAGA GGCCTCAATA
18651 AGGCCAGTGT GTTGACTTTG TCCTGCAGTG CTCAGCAGTG GAAAAAACCA
18701 ACAGCCACGC AGGGAGAGGG AAGGAGCCAC GATGGGCACG GGTTACTGGG
18751 GCCAGGGCTT GACTGGTAGG TGGACACAGC TGAAGGCCCA GGTTGTGTGG
18801 GAACAGAGCG CAGAAGCAAT AGATTCCTCT TGAAGATCCT TGGGCTGTTA
18851 ACCTTTTTTA AATTTAAGAG AGGTTGTGTG GGCGGGGAGG GAGGAAGGAA
```

FIGURE 3, page 6 of 14

```
18901 AATCCTTCAG AAGACATAGA CTTACTCTGT TTCTTCCATC ATATGTGAAT
18951 GCATATGAAT AGCCAAAAGG TGAATAAAAC ACATGTTCCC AGGTGGCCAG
19001 TGAGACCTAG GTTGCAAGAT GGTGGGGTGT GTGTGAGGCC GGGGAGTGCT
19051 GCGAGCCCCG GAATTCCTCA GCCTTAGTCC CCCGCCACAT AGCTAAGAAG
19101 TGAGGGAGGA GGTGAGAAGG AGTCACTGCC CAGCCTCACT TCCGGTGGAG
19151 TACCCTGTCT CCTTGTCAGT TCTGTCTCTG GGGACAGTTG CCTGCTTTCA
19201 CCTCTCCCTC CATCCCCTCT TCTCTCACAG GGAAAAATTC ACCTTAATAT
19251 TGGAAGTTCC TCTCCTAGCA AAGTCCTTCT CAGGCACCCA CAGGCAAAAA
19301 GGAAACTAAG CAGAGTTAGG GCTTCCAGGC CTAGCCAACT ACACGACTCT
19351 CCTCTTGCTT CCCTAAGAAC CAGCGCAAGG GGCAGCGTGG GTTCCAGCAT
19401 AGATGGACCT GTGTTGGAAT CTCTGCACGT GCTGTGCTGA CCCTGGCTAG
19451 CCATTGACCT CTCTGAGCCC TTGTTTCCTT TCCACTAGGC TCTCTGAGGG
19501 CAGGGGCCAT GTCTTTTTCA CTGCTCTGTC TGCACTGAGC ACTGTGCAGG
19551 GCACATAGGA AGTTCCCATA AATGTTTGTG GGATAAAGGA AATAAAACCT
19601 TCTCTCTTCC TGTCCCCCTT GTGATGGCTT TGCACAAGGC ACTGTCCTTG
19651 GCCAGGTTTG CTAGGCTAGT GTGAGGATAA ACCAGGTATA TTACAAATTG
19701 GAGAAAATTT CTCGTTCTTC TTGGAAGAAG GTGCTGTATC ATGAAACAAG
19751 AATGTCTTGA TTCCCTTCTA TGCCAGGTAC TGGGGAGAAA CAGGTGCCTG
19801 ATAACCGTTG ATCCAGGCAG AAATAAGCAT ACTCCTGCTT CCCAAGGCCT
19851 GATGCTTCTC TCCTTCCTCC CTTCCTCCCT CCTTCTCTTC ACTCTTTCTC
19901 TGCACACATG GAAGAATGGC TGCCAGGCAT TGCCCATTTG GAAAAGTACA
19951 GCTCAATGGA TATGAATCAG CTTGGGCAGG CGAGAAATGA TTCACGTCTG
20001 ACCAAATCGA TTTAGTTCAG GTTGCCCGTT CTGCATCTTT TTTCCCTTGT
20051 AATTAAATGA TGATTGGTCT TGATGGTGGG AAGGAAGAGA CAGAATTTAA
20101 TTTGTTTGCC TTTGTAGAAA GCTGGGGACA GCACAGATAA GGGAAGATGT
20151 CTCCCATTTG GCAAATAACT GATGCGGAGG TGGAGTGGCA GTGGTGATGG
20201 GGATGCTGGT GCCTTCAGGC CTTCTGGGCC GGGCAGTGCA GCTGGTGGCA
20251 GACGGTTCGG AACTCTACCA TGTTCCCATC TGAAAACTGT GGCTGATCAT
20301 GCCCACTCCT GACCTTGCTC CAGGGAGTAC ACAAAGACGT AAGCTTAATT
20351 AACCCACCAG ACGTAGCTCT TGAATCCCTG GGCATAGTGC CTGGGTATAG
20401 TTAGAGTTGG GGAGAGGCAT GGTCAGCAAA ACAACCTCCC TCATCTCTCT
20451 GTTGTCACTC AGAGTCAAGC TGGCTGCTGC TGGTGGTGCT GACTTCTCTT
20501 GCTGCAGATT TCTCCAATAT GTTTCTGCCC TGCACGCATT TGCCAAATCC
20551 CTTCGGTTTC TTGTGTCTCG TGGCAGCTTA GCTCCTCCAG CCCTTGGATG
20601 AAGAAGCGTG GGAACTCTTT GCTTCCTTTC CCTCCCGCAG TGACATGCCA
20651 TGCCATGCCA CTGCCTCTTC ATCTGGTCCT ATGACAGTCA CTCATAAGCA
20701 CCCGCATGTA CCCGGCCCTG CACTAGCTCA TGACAGCTGC AGTCAATTGG
20751 GCCAGGTGCT GTATCTCATC CGGCCTCCTC AGCAACCCTC TGAGATACTG
20801 GTAATGTCCC TGATGAAGAT ATTTACTGAG GCAGAAATGG ACGCTCAGTG
20851 AAGCAAGGTG CCTGATGTTA TAGCAATGAG CTATGAGTGG CCAGAGGGAG
20901 GAGATAAGCT CAGGCCTGAC ACCAAAGCCC ATGCTCCTTC TAGTCAACCA
20951 CAGTGCCTCC TATGGTGAAT GAGTGAGTCA GCAACCAAGA CGCATGAGGC
21001 CTTCTTTTTG GTGAGCCTTG GCTGGGTGCT GAGGCTTCAG GTACAATCAT
21051 GGGTTGGAAG AGCCCTCCTC TCTCTCCACA GTCTGGCACT ATGACCCCTT
21101 CTGGTTATTA ACAAGGCAAA GAGAGAGAGG GAAGAAAGCA GGCAAATAAT
21151 GTGGGTTGCT ATTCCTAGAG ATTAGAATTT CAGGAAGGAT AAACACAGCG
21201 TTCTCTCCAG AAGTATAAAT AGGAAGACTT CACACATGAC TAGAACGAGA
21251 CATGTTTTAA GTCTGTCGAG TAAGGCAGTG ATGAAGTAGA TTTCCCCAGA
21301 TTCACTCTCC CTCCTCTGGG TCCCCCAGGG CCTTTACTTG TGGCAACTTT
21351 CAGCTCAGGG AGGGAGGAAA GCCCCTTTCA AAGCTTCAGA TACTTCCTTA
21401 AGGTCAGTTT CTGCTTAAAG AAGGCCTTTA CATTACTTCA TCCCTTTGCC
21451 AAATTAAACT GAAAGGAAAC CTTTCAAGTG TGATTGCCTG GCCCTTTCCT
21501 GTTCATTTCT CGTGGGTACG CTTTCTAACT TTCTTTCTTT CTTCCTTTCT
21551 TCAGGTGTTG ACTTTAAGAT GAAGACCATA GAGGTAGACG GCATCAAAGT
21601 GCGGATACAG ATCTGGTGAG CTGGGGAGGA GGAGGAGGCA GATGTAGGAG
21651 AAGAGGACTT CTGGCTGCTC CTTAGCTGCC CCTGCCATGT GTAAAATTCC
21701 TAGGCTTCAC CTGGGATAAC TGGCCACCTC TCTGATGGAT GGAAGCGAAG
21751 TCTCAGAAGC CCATCTCTTC CTATAAGCCT TAATCTCCAA CCTCTAAGAA
21801 ACTTTAGGGG ATTGACTACA AGCACCAAAG GGCAGGAATT AGAAGGAACT
21851 GGCACACTAA CCATTGTGAA TTTATCTCAG GATTAGGCTT TGCCCTTGGG
21901 CTGTGCCACA CTATGTTAAG ATTGGAAGGA AGGAGGCTAC ACCCCCCATC
21951 ATTTAGGGCG AGACCCTGAG AGAGTTCCTC AGGATAGCAT GATGAAGTTT
22001 CCACAGTAGC AGAGGGTGCT GCTGTGGCTC TCTGCCTGAG GTCTTGGAAG
```

FIGURE 3, page 7 of 14

```
22051 CACTGCCTTT GCCAGGGTTT AGAGCTCCCT CTCAATTCCA CAGCAGTATG
22101 GGCACTGCCT TCAGAGGTCC CATAGGGACT AGGGGTGTAG CAGCATCCCC
22151 TGCCAACTCC CATCCAACCA AATCTGGCCA CAGTGGCCAG ATTCCAGAGA
22201 GCTGTCCAAG GCCTGTTCTG GCTGTGGCTT CTGGTTTCTG CCAGGAGGGC
22251 AGTTGGCAGG AGGGGCCAAG GCCCTGCAGG CCTGGTCAGC ACCAGCACAG
22301 ATGACCAGGC CTCTGACTGC AGATCCCTGT GGGGATCCAA GCATCCCTGG
22351 TTTTTCACCC TTTAGCTCCC CAGTTTTTCC TACAAGGGGA CAGCTCTGCT
22401 CTTCCCCTCC CCGTCTGTTC CCATGGTCCC TGCTCCTCTG AGGGACTGGC
22451 TTTCTCCTGC AGGGACACTG CAGGGCAGGA GAGATACCAG ACCATCACAA
22501 AGCAGTACTA TCGGCGGGCC CAGGTAAGCC ACCACATTGG GGGTTTCAAA
22551 GTGGGAAGCT GCCACCCACA CTCCCAGCTC TGGGTATTTG AGATGTCTGT
22601 GCCACGGATC CCCTAAATAC AGTTCGCCTG CTTGGAGGAG CGCAGGGCGT
22651 CTTTCAGCTG TTCACTGATC ATTTGTCCGT CCATTGTTCA TGGCCCACTC
22701 ACTGCAGGCA GGCCCCTGCC CTCACCCCTG ACTTCCACCC TCCATCCTGG
22751 GTCAAAGATC CAGGTCAAAG CATGTGGTGT CTTCCTGCTG TAGAGAGTTC
22801 TGTGATGGGC CTGGGAGGCG GCAGTGGTGG GGTCTGAGAG AAGAGATATT
22851 TCTGGATGCT GAGCAGGGAG AATGGGAGAG TGGGACCCAA CCTTTAAGTT
22901 TCCACGGCCC CTTCTGGCCC CATGACTGCA CTCTCTCTGT GCATATCACA
22951 TCTCTCTATT TCTCTCTCTC TCAGGGGATA TTTTTGGTCT ATGACATTAG
23001 CAGCGAGCGC TCTTACCAGC ACATCATGAA GTGGGTCAGT GACGTGGATG
23051 AGGTAGGAGA TGCCACCTCA CTGCCGGGGT GTGGAGAGGG TGCCTCACCG
23101 GGGAAGGCAA GGCGAGGGCC AGATGGGAAG GCAAATGCTT CCAGGAAGCT
23151 TTGCCTTCCA CAGCCCTGGA TGAAGACCTC TGGGTGAGTA AGACATGGGG
23201 AAGAAACCGA AGCTGCCATG CCCTCACTCT CTATACCCTG CCAGGCCTCC
23251 ACGGCTGTGT CTTTCCCGGA AATGAATTAG TTCCAAGTCT TCCCTGTGAG
23301 CAGCTTCTTT CCTGAAATCT TGGGACCAGG TGGAGTTGCA AGATTGGGAT
23351 CTAGTCCTGG CTCTGCACAA TAGCTGTGGA GCCTTGGGAA GCCATTTGAA
23401 TCCTCTGGGT CCCCAGTTCC TGTAGAATGA GGGCTGGACT TACATCCAAT
23451 GTCCTTTCCA GCTCTGATAC CAGTGGTCTA ACCCAAGGAA GCACCAGTCT
23501 TAGCCAGAGT GTCTTCTACC CTAAGCTCTC CCCGTGATAC CCTTGAGGTC
23551 AGCCATGGCA CTTGGGGGAG CCTGGCACCT GCATCCAGTC GGCCCACCCT
23601 GTCCCTAGGG CTCTGGAATT GGTGGTGGGC TGGAGGCAGT GCAGACTCTG
23651 TAGGGAAAAT TGGGGGGGCA GGCAGCACTC ACTGGCTGTT CTGCCCATCC
23701 TTTGTCCCTA GTACGCACCA GAAGGCGTCC AGAAGATCCT TATTGGGAAT
23751 AAGGCTGATG AGGAGCAGAA ACGGCAGGTG GGAAGAGAGC AAGGGCAGCA
23801 GGTAAGTGGA GGGAAAAGGC AAGTCCACCC CAGGTCCTCT GCTGGGCCTC
23851 CAGGGCCAGT CCTGAGCGTG GGGACCTAGG GGTGTGTTCC CCAGTGGCAG
23901 GTCCTCCCAC ACGTCCCCAG CACCCCAAGG CCCTGGGGGA GTGGCCATCC
23951 TCGGAAGGCT TGTTGTCTGG GTTTCAGGAC AGAAGCCCAG AGATTCGGGG
24001 TCCATCCAGA AACAAAGACG TCATAGGCAG CAACTCTCCC AAGTCCAGGT
24051 CCCCAAATGC AGGATTGCCC TCTGCTTAAG AGATCATCCC CGTGTTAGTA
24101 ATGAAGGACT TCAAGTTGTC AACCTCTTCT CTGACAGCAT CCAGGCCTAG
24151 CTGCCATGTT ACGGTCGAGA AATGATCTCC CATCCCACCC AACACTCCCC
24201 CACTCCTGTC CTTCTTACCC AGGAAAGAGC CAGGGAGGCA AATGAGGAGA
24251 CAAAGAGCCA CAGCTGGAGA AGCCATGGGG GCAGAAAGGG TAGGAGGATG
24301 ACGCTGAGGG AATGTCCAAG CATGCAGGGA GACCATCCTC CCAGAGAGCA
24351 GAAAGAAATA TTGGTTATTT TTTTTTTCTT TCTTTCTTTT TTTTTTTTTT
24401 TTTGAGATGG AGTCTCGCTC TGTCACCCAG GCTAGAGTGC AGTGGCGCCA
24451 TCTCGGCTCA CTGCAACCTC TGCCTCCTGA GTTCAAGCAA TTCTTCTGCC
24501 TCAGCCTCCC AAGTAGCTGA GATTACAGGT GCATGCCACC ACGCCTGGCT
24551 AATTTTTTTG TATTTTTAGT AGAGATGGGG TTTTGCCATA TTGGCCAGGC
24601 CGGTCTCGAA CTCCTAACCT CAGGTGATCC ACCTGCCTCA GTCTCCCAAA
24651 GTGCTGGGAT TACAGGCGTG AGCCACTGTG CCCAGCCAAG ATTGGTATTT
24701 CTGAGATAAG TTATCCACTC AGTCCGTGGA CCTCAAGAGT TTTCCTCTCC
24751 CTTTTCAGTC AATAGCGTTC CATTAGTACT TAAAATGAAA TTGATTGTTT
24801 GGTATAAAAT ATAAGACATG GTCATTGACC AATTTGAAAG TAGAGGCAAA
24851 GCCTACTAGG ATAGTATTTA TTGAGCACTC TATGTGTGGC ACTGTGCTAA
24901 GGCAAGCGCT TTTAAGTGCA CGACCCCACT GAATCATCCC ACAACCATGG
24951 ATGGGAGACA CACTCAGTCT CCTTTAACAG AAGATAAAGC TGGGGCTTAC
25001 AGAGAATGTA CAACTTGTCC AAGGTCACAC AGCTAGCCAT CAGTGGCAGT
25051 GCTGCTATTC AGGTCTGGGA CTGTGGGACT CCAGAGCCCA TGTTTTTTAC
25101 GAGGATGCCA TACTGCCACA ATGGATGGTG TCTTTATCTC CTGATATATG
25151 ATTGTGTGTT GGGAGGCGTG GGGTGGCAGC TGGAAGAATG GAGAGGCATA
```

```
25201 TTTGTGGAGG ATCTTCCCCC ATTCTCTGCT ACCCTCTCTT GGAGCTCCCA
25251 GTCCCATCTG AGAAATTATC TACTCTGAGA AATCGTCACA ACACAGCATG
25301 GTTGTGAGTG CAGTGGCAGA AGCCTGTGCC TGGTTGTATG GGCCCCTCCC
25351 CTGCCTTACT GACTCTCTTT CAGAAATGTC CTTCTCTTGC AGCTGGCGAA
25401 GGAGTATGGC ATGGACTTCT ATGAAACAAG TGCCTGCACC AACCTCAACA
25451 TTAAAGAGGT GAGAGCCCTG GTGACCAGGC GCCCGCTCTC TCGGGCTGAG
25501 TCCAGCAGAG GTGGGAGGAG GAGCCATAAG ATGGACCTTA TCCCTCAGGC
25551 CGCTGCAGGG TTGCCAGGGG AGAGGAGGAG ACACTGGACT AACCTGTGCC
25601 CTTTGGTTTC CAGTCATTCA CGCGTCTGAC AGAGCTGGTG CTGCAGGCCC
25651 ATAGGAAGGA GCTGGAAGGC CTCCGGATGC GTGCCAGCAA TGAGTTGGCA
25701 CTGGCAGAGC TGGAGGAGGA GGAGGGCAAA CCCGAGGGCC CAGCGAACTC
25751 TTCGAAAACC TGCTGGTGCT GAGTCCTGTG TGGGGCACCC CACACGACAC
25801 CCCTCTTCCC TCAGGAGGCC CGTGGGCAGA CAGGGGAGCC GGGGCTTTGC
25851 CCTGCTGCTG TCCTCTCGTG TGATGACCCT ATTGAGTATC AGTAGCCACT
25901 ACTCCCCCTG CCTGGCCCTG AGAGCGGCTC TGCTGTCATC TCAAGCAGCC
25951 CCTGTCCCCA GCCCGTCCAC CCTGGAGTGG TCTTCTTCAG CCTGTTTCCC
26001 CAGCCACAGG CCTGCTACGA CCCCCACGAT GTGCCGCAAG CACTGTCTCA
26051 CCATCCCGCA CCCACCAGAC AACAGCCAGG GCTGGAGTCC AGGCCACTTT
26101 CAGCTGCTCC TTTCTCCGTG CATCGTGTCT CTTCTCTGCT TTTTCTCTCT
26151 TCCCCCACTT CTCTTTCTCT GACCCCTCCC CTCCGGTGCG TTTCGTATCA
26201 AAGCTCCTCA AACCCCGTCC CCCGTGTGTC CTGCTGTGTG CAGCTCGCTC
26251 TTTCCTTCCT TCCTAAGCTA TCCAAGGGGA TGGACCCAGG CTCGTGGGGA
26301 GGTTCCACCC TTGGATCCAG GAAGAACCCT CCACCCTGCC TCGTGGGTGG
26351 GCCAAAGGCT ACAGGGTGCT TCTTCCTCTT CCCCCACCCC CACTGTCCCT
26401 CATGTGCCAT GGGCCTGCCT CCCCAGTGAC CTGCGAAAGT GGAGCATCGA
26451 GGTAGGAGGG AAACGGCAAC CAGGGAGTCC TCGAGCCTGG GGCTGCCCTA
26501 CCTCTACCCA TTCCCCGACC AGAGCTTTGC CCTTGCTTGG CTGCCCGCCT
26551 GCCTCTTTGG GGAACTGAGC TCAGAGGCAG GTGCTTCAGA GAAGGAAACA
26601 AAATGAGGGG TGGCAGGGAT AAAAAGTCAC CTCCATTCTC TACCTCCCAT
26651 GCAGCATGAA CACAATTTCT CTCCACCTGG CTCCCAAATT TAAAGATGTG
26701 GACCAAGGCC TGTGGGTACT CCAGGGGCAA GGAGAGCCCT GGGGTCAGTG
26751 ACACTGTCAG GCCAACCATG CACTCCACAA AGGGGAGCAT TTGGAAATGA
26801 AGGACTAGCT CCTATGTATC AGGTTAAGAG CAAGGGAGAG CTGGCCAGGG
26851 ACAGCAGTTT GCACAGCAGA GGGGAATGTA GCAACAGCAG GGCCTCCTAG
26901 GCCCCATCTT CCATTTCTTA GGTAAGAAGA GCATTTCCTC AGACTCCCAG
26951 GCGGAGGACT GAGCCTAGCC TTCAGCAACC AAGGTTCTCC TGGGACCCAA
27001 AGTTTATGGG AGAAGGGCAA AGACTTCATG GGAAGAGAGA AGGAAGGCCC
27051 TGGGTAGAAA CGCTTGGTGC TGTTCTCTTT GGCCTTTAAG ACAAAGCGCT
27101 CATCTTGCCC TCTACCTCCT GATAGGCTTG AGGGTTTGCC AACCACACTG
27151 TGGCTACAGG TGGAGGGAAG AGGACTCCTT CCTCCAGAGT GCTATGTTCA
27201 GGAAGTTTCT TTAACCCCAT ATGGCCCAAG AGTAGCTCGT AGGAGGCCCT
27251 TTAAAGACGG AACAAGTAAT TTACCAGTTC TACTGGGGTT CCTGCCCACC
27301 GTCCCAAGGT GGGCGAGGCC TAGGAAGAGG GTCATTCTTA AGCCACACAT
27351 TAGCTGCACT GCGTGGCTGC AGCCAAAACA AAGAACTGGG TGTTGAGTAT
27401 TCATCAACTA AGAACCAAAA TCCAGGGCAC TCATATGTGA AGGATAAGAA
27451 CCTCACTTCC TTACTCCTCC AAAAAGAAGT GGGGAAAGAA CCATCAAACC
27501 TTTCCTCCTG ACTTACCAAA CCAGGAAAAC AGCAGGAGAG GGTGGCTCAG
27551 GACTTAGGGA CAGGGTATAG CTTAGATGGT GGAAAGCAAA GGAGAGCAGG
27601 AAGTTGTAAA TCACTGGCTA ATGAGAAAAG GAGACAGCTA ACTCTAGGAT
27651 GAAGCTGTGA CTAGGCTGGA GTTGCTTCCT TGAAGATGGG ACTCCTTGGG
27701 TATCAAGACC TATGCCACAT CACACTGGGG CTAGGGAAGT AGGTGATGCC
27751 AGCCCTCAAG TCTGTCTTCA GCCAGGGACT TGAGAAGTTA TATTGGGCAG
27801 TGGCTCCAAT CTGTGGACCA GTATTTCAGC TTTCCCTGAA GATCAGGCAG
27851 GGTGCCATTC ATTGTCTTTC TCTCCTAGCC CCCTCAGGAA AGAAGGACTA
27901 TATTTGTACT GTACCCTAGG GGTTCTGGAA GGGAAAACAT GGAATCAGGA
27951 TTCTATAGAC TGATAGGCCC TATCCACAAG GGCCATGACT GGGAAAAGGT
28001 ATGGGAGCAG AAGGAGAATT GGGATTTTAG GGTGCAGCTA CGCTCACCCT
28051 AAACTTTTGG TGGCCTGGGG CATGTCTTGA GGCCCAGACT GTTAACCAGG
28101 CTCTGCTGGC CTGTTTACTC GTCACCACCT CTGCACCTGC TGTCTTGAGA
28151 CTCCATCCAG CCCCAGGCAC GCCACCTGCT CCTGAGCCTC CACTATCTCC
28201 CTGTGACGGG TGAACTTCGT GTACTGTGTC TCGGGTCCAT ATATGAATTG
28251 TGAGCAGGGT TCATCTATTT TAAACACAGA TGTTTACAAA ATAAAGATTA
28301 TTTCAAACCA CCGGTGTGGC TGCCTGGATG AGTCCTTGGG GGTAGGTCTC
```

```
28351 ACTCAGACCC TGGCAGTGAT GTGGGAGGGA GAGAGGCAGT GCTGGTAGAA
28401 GCAGCTCCAG AAGCAAAGGC AACAGCAGTA GAGTGACCAC GGAAGCGGCA
28451 AACATTGTCT TCCCTTCTCT ACCTTCCCTA GTGCCACCTG CAGGGAGGCC
28501 CAAAGCAAAG CCCCGTTGCC CTGCATTGGG CTGGCACTGC AGAAATAAGA
28551 TGAAACACAG TTATCGAGAG GATGCTGAAC ATCTATGAGC AGGTTTTAAA
28601 GCCAAGATGA GTCTCATCTG TTTGTGTGGG TCAGGAACGG GTCTTCCTGA
28651 AGGCATGAGG TGGGACTGGA TAATCTTTCA GATTTGTGAT TGGATACCTC
28701 GGGGGAGCAG AGGCAGACTG GGATCTCAGG ACTGCAGGTA TTTCATACTT
28751 TGGGATATGG AATTGATGGA  (SEQ ID NO:3)
```

**FEATURES:**

| | |
|---|---|
| Start: | 2044 |
| Exon: | 2044-2167 |
| Intron: | 2168-21554 |
| Exon: | 21555-21615 |
| Intron: | 21616-22462 |
| Exon: | 22463-22523 |
| Intron: | 22524-22974 |
| Exon: | 22975-23052 |
| Intron: | 23053-23711 |
| Exon: | 23712-23801 |
| Intron: | 23802-25392 |
| Exon: | 25393-25458 |
| Intron: | 25459-25613 |
| Exon: | 25614-25769 |
| Stop: | 25770 |

**CHROMOSOME MAP POSITION:**
Chromosome 14

**ALLELIC VARIANTS (SNPs):**
DNA

| Position | Major | Minor | Domain | |
|---|---|---|---|---|
| 206 | - | T | Beyond | ORF(5') |
| 4963 | C | T | Intron | |
| 8175 | G | A | Intron | |
| 10515 | T | C | Intron | |
| 13034 | T | C | Intron | |
| 13781 | T | C | Intron | |
| 14050 | A | C | Intron | |
| 14273 | - | T | Intron | |
| 17582 | T | C | Intron | |
| 17700 | C | T | Intron | |
| 18074 | T | C A | Intron | |
| 19328 | G | T | Intron | |
| 19570 | A | G | Intron | |
| 20892 | C | T | Intron | |
| 26465 | G | A | Beyond ORF(3') | |
| 26472 | A | G | Beyond ORF(3') | |
| 28071 | C | T | Beyond ORF(3') | |
| 28096 | C | G | Beyond ORF(3') | |
| 28403 | A | G | Beyond ORF(3') | |
| 28467 | C | G | Beyond ORF(3') | |

Context:

DNA
Position
206     GCTCAAGATTGCACAGCTGGTGAGTGGTGACACTGGGACTGGAACCCAAGTGTGCCTTAC
        TCCAGAGCCCTTGGCATGCACCTGAAACCCCATGTAAGCCCACTGTGGAGACGCGCACCT
        CGAAATAATGGAATCCACTACATCAGTTCCTTTAGCTTTCTGTGTAATCAGAGTAGCTAG
        CAGGCTCGGGATTTCGCCCCCCGGC
        [-,T]
        TTTTTTTTTTTTTTTTTTTTGAGACAGAGTTTTGCTCTTGTTGCCCAGGCTGGAGTGCAAT
        GGCGCAATCTCGGCTCACCGCAACCTTCGCCTCTCAGGTTCAAGCAATTCTCCTGCCTCA
        GCCTCCCGAGTAGCTGGGATTACAGGCACCGGCCACCACGCCCAGCTAATTTTTTTATAT
        TTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTTTTCCCCTCTT
        ATTATAATTCAGACACTTAACCTGAAATATACCTTTTCAAATGAAGTAAATGGGCTTACC

4963    TATTAAGGGACTTGGGATTCTCCCTTATCTTGGGCGTGTTTTTCAGCATTAACTAAAACT
        TAAAGGAAAGAGTTGGATGGTCAAGAAAAGCTTTTTCCTTAAGTGATATGGACAGTTTCT
        CAAGGAGGTAGAAGGGGCAGCCAGGAGACAAATCAAGGAGCCAACGAAATGAGTGCTACC
        AAGTCATAGTCATTCGCTTATTTTTAAAAAATGCGTGTCCTGTATGCCAGGCTCTGCACT
        GAGACCGAGAGATTCCAAGATGAATAATACCTACAGTCACTGTTCTCAAATTGTGCATTA
        [C,T]
        CTAAAACACATTACATGACCATGCTGGCCACTGATCGAGGCACCTTTCCCAGGGGCTTTT
        TTTGTGAATTAAGAAAACAAGGTAATTCACCAGTTATTGCCAAGATAGTTTGGCTTCTTG
        GCTCATGTGGATATCACCTAGGCCAGTACTTTTGTGATTTACTGTGTACTCCACTTTAAC
        GGCCTGCGATCTTCTAGAGAAGAACCCGCCAGGGAGCAGTGAGAGGCCTCCCTGGTAGAC
        TGAGACACTGACTGTCCCTCCCCCTATCCTTTTCGTCTTTCTGGCCAGCAGACCAGCAGG

8175    ATGCCAGGTGCCATGCTAAGATTTGGGGACACAGTGGTGACCAAAACAGACAGAAACCAA
        GGAGCTGGCTTACATTCCAAGGGAGTGCATAGGAAGCTGTGTTTCATTTCAGTTTCTGCT
        CTAGTACCCCCCTTTCCCTGGCAGTGCCAGGGTCTGAGAAGGAAGAGTGAGGTGGTGAGG
        AGGTGTGAAGCAGTGGGGTGACCTGAGAGGAGAGGATGGGGTGGCTTTGCCTCAAGGCTT
        GGGCCCCTGCTAGGTGTCGCTCTGCCTCAGGCCTCTGTTTCTCCTCCTGACACAGGCACA
        [G,A]
        ACTCGGCCTCCCACCCCTTCCCCAAGGACATGACCTTGGGAAGGAACATATCTGAAGCCC
        GCGGAGGGTTTCCGCTGCTGTGCATCTGTGCCACAGATCCGCAGATGCACCCACAGCTGG
        GAGCACCGGTTCCTCCCGCCTACCTGCACTCCCTGGTTTCTGTTCCTTCCTCCTCCTCCT
        TCCTTCTCCCCGCTCCCCAGACAGGCTGGTGATGAGCTTTATAACATGAAAGCTGATATT
        TGGCCATTATCCTTCTACCCTGATTGCCAGCTCTTCTCAGAGTGCCTTCTTCTGTAATCC

10515   CTGGTGAAGGCTTTGAAGAGGAAGTGACATTTGAGTGGAGTCTTGAAGACTAGGCAGGAT
        TCTCCAGGGGCCCTGGGTGTGGGGGAAGCACACATCCTCTTCCCTGTAGGAGGTGCTGTG
        GAGAACACCTCCAGTGGGGCTGCTACTCTTCAGCCTTGCTGGGGCCAGCTGGAGTGGCCA
        CACCATGGTCACACCAGCTGAAGTTCAAGAAGCCCCTTGCCAGGAGATTGCTTTGCTGGC
        TCTGGGTGAGGGCAGGTGCATCTGGAAGCCCCCTTCTTTCTAAGATGTTTGCTCCTGAGT
        [T,C]
        TCTATGTCCTAGTCTTTTCTTCCCTGAACCTTTTGCTACCAGTCAGCACAGCCCTGCCTG
        AGAAGGAGGCTGGAGGAGTGAGTGGTCAGTAGCCTGGTGGGTCTTGGCTGCCTCTGTGGT
        GCCCGCTGGCCTAAGTAGCAGGCTTAGGGAGGCGAGACCCAGTTCCAGGGGCTGCCAATG
        GGGAGCGAGATGGGGTGGCTGGAGCACACTGCACATGTCACCAAGGCTCTAGGGAGGTCT
        GTGCACAAGGCAGTGGGAAAAGCAAGGGGAAGACCCAGCCTGGTCAACATGGTGAAACCC

13034   AGATTTGGGTGAGGACACAGCCAAACCATATCAGCTCCCGGGATCCCTGTGTGAATGGGG
        TCTTTTTTGGTGTTTGAGGGCTGCACAGGGTGACCTCTTTAGAGGTGACCTCCTGCCACA
        ACCCACAGGAGGTGCACATGGCCCACACATGCTGGTTTCCTGCAGTGGGAGGGGCTGGGG
        CACTCCTGGGACCTGTGCTTGGTAACTGGAGCTGGCCTGGCCCTGGGGATTGGGTGTCTG
        CCTTGGGTTTCAGGTGTATTAGGTTGTTCCTCGTTGTGGAGTCTCATTACTAATGAAAAG
        [T,C]
        TCAGGTCGCACTGCTGGTCCTTTGGGCTGTGGTTGATCCTGGTGATAACATTTGGCACCC
        AGAGGCAGCCCTGTTTCCACTGAAGCATGCGGAGCTTGGCTGGCAGGCAGGCAAGCTGGC
        AGCTGCCCTTAACCCATGAGGTGCTGGCCCGCTAGTAGGCACACCCTACCTGTGCCAGAA
        TTGAGGTTGTAGCCAGACTCCAGGAGCCATCTGGGCCCCACAGGGGGCGGCATTTCCTCT
        TTTTGTTGAAACATTCCAGCCAAGTGCTGGCTTGGGCTTCATCTCTCTGTCCCACTCTCC

FIGURE 3, page 11 of 14

13781    CCCTGTGTTATGGGTTTTACACCTTATCTCACAATCTTAAAAAAAAAATTCTCTGAGAAT
         CCTCTGTCACCCCCACTTTACAGGTGAGGAAACTGAGGCAAAGATAGGCTAACTGGCTTC
         CCCAACACCATGCAGGTAATTAGTGATAAAGGCAGGGTTGGAACCAAACTTGACCTCCCA
         ATTGTGCTCTTAATGGCCAGGACACTCTGTGTCTTGAGCCACACTTCCTCCATGTTTTCT
         AGGGCTTTCTAGGGAGGCAGACAGTGATGGGAAGGGGTGTTCTTTAGTGTGGATGTGCCC
         [T,C]
         GCCTGCTCCTTTCTGTAAGCGTCACAGCACCTCCACTGCTGTACTGGGGAGGCACCAAGT
         TTTTCCCTGTTTGCCCACCCAAGGCGAGCTAGCTTAGGAGTCACGTGAGTGCTGGGTGTC
         TCGCCTGCTGCATCCCTCTATCCTGCCCCTGCCCCCGGTGCCCAGAGGAGGGCCCTGCCT
         GTCTTCCCAGTTCTCCAACAGCAGCGCTGTCCCAGCACCCTCGGGCTCCAGTTGTGGCCT
         GGCAGCTGCTGGGGCAGACACCATACAGACAGAGTCACAGCAGGAAGAGGATGGGGCCCA

14050    GGAAGGGGTGTTCTTTAGTGTGGATGTGCCCTGCCTGCTCCTTTCTGTAAGCGTCACAGC
         ACCTCCACTGCTGTACTGGGGAGGCACCAAGTTTTTCCCTGTTTGCCCACCCAAGGCGAG
         CTAGCTTAGGAGTCACGTGAGTGCTGGGTGTCTCGCCTGCTGCATCCCTCTATCCTGCCC
         CTGCCCCCGGTGCCCAGAGGAGGGCCCTGCCTGTCTTCCCAGTTCTCCAACAGCAGCGCT
         GTCCCAGCACCCTCGGGCTCCAGTTGTGGCCTGGCAGCTGCTGGGGCAGACACCATACAG
         [A,C]
         CAGAGTCACAGCAGGAAGAGGATGGGGCCCAGGGCTGCTGCCTCAGGCCATGGCTGCATG
         GCACCATCAGTTGATTGAGGAGCTTTTCTTGCCAATGTCTGAGGCATCAGGTGGCAGGAC
         ACGTCTCCCTGCTCTTAAGCCTCAGGCATGCAGCCCTTCTTATGCTCTCTGGGGTGAGGG
         GGAGATCCCCCTCATGGAATTGCTTTTTTTTTTTTTTTTTTTTTTGAGACAGGGTCCT
         GCTCTGTCACTCAGGCTGGAGTGCAGCCTCAACCTCCCAGACTCAAGTGATCCTCCTGCC

14273    TCTCCAACAGCAGCGCTGTCCCAGCACCCTCGGGCTCCAGTTGTGGCCTGGCAGCTGCTG
         GGGCAGACACCATACAGACAGAGTCACAGCAGGAAGAGGATGGGGCCCAGGGCTGCTGCC
         TCAGGCCATGGCTGCATGGCACCATCAGTTGATTGAGGAGCTTTTCTTGCCAATGTCTGA
         GGCATCAGGTGGCAGGACACGTCTCCCTGCTCTTAAGCCTCAGGCATGCAGCCCTTCTTA
         TGCTCTCTGGGGTGAGGGGGAGATCCCCCTCATGGAATTGCTTTTTTTTTTTTTTTTTTT
         [-,T]
         TTTTGAGACAGGGTCCTGCTCTGTCACTCAGGCTGGAGTGCAGCCTCAACCTCCCAGACT
         CAAGTGATCCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACCACAGGTGGACACCATCACA
         CCTGGGTTTTTTTGTTTTTTGTTTTTTGTTTTCTAGAGATGGGGTCTCACTTTCTTGCTC
         AGTCTGGTCTCGAACTCCTGGGCGCAAGCAGTCCTCCCACCTCGTCTTCCCAAAGTGTTT
         GGATTACAGGTGTGAGCCACTGTGCTTGGCCTTTTTATTTATTTAGAATTTGTTTTGGAA

17582    GGATGTTTCTTCCATGACATATATAGCTCTTGAAACTACTTCTATCTAATATCACCCACA
         GTGCTGTTAAAAATACAGATTTCTGGGCCTCACCCTCAAATTATGATTCAGTAGGTCTAG
         GCACGTCAAGGTCATTGTTTTTGTCTTTGTTTTAAGTCACCCCAGGTGATTCTAAAGCCG
         AAGCTCTGCAAAGCACACCTTGAGAAACAGAGAACTCTTGTGCTCTCGCTCTCTTGACAC
         TTCAGGTGCAAAACTTTTGTCCTAATGTCGTTCTCAAACTTACGCATGTGTGAGAATCAC
         [T,C]
         GTGAGAGCTTATTGAAACTGATTGCGGGACCCCATACCTAGAGGGCCTGATTCTATAGGT
         CTGAGGTAAGGCCCAAGAATTTGCATATTTGCATTTCGTTTTCTTTTCCTTTCTTTTCTT
         TTTTTTTTTTTTTGAGATGAAGTCTCACCCTGTCGCCCAGACTGGAGTGCAGTGGCATGA
         TCTCAGCTCACTGCAGCCTCTGCCTCCTGGGTTAAAGCGATTCTCCCCACACCCCAGACC
         CGCTCCTGAGTAGCTGGGATTACAGGTGCCCGCCACCATGACTAGCTAACGTTTGTATTT

17700    AGGCACGTCAAGGTCATTGTTTTTGTCTTTGTTTTAAGTCACCCCAGGTGATTCTAAAGC
         CGAAGCTCTGCAAAGCACACCTTGAGAAACAGAGAACTCTTGTGCTCTCGCTCTCTTGAC
         ACTTCAGGTGCAAAACTTTTGTCCTAATGTCGTTCTCAAACTTACGCATGTGTGAGAATC
         ACTGTGAGAGCTTATTGAAACTGATTGCGGGACCCCATACCTAGAGGGCCTGATTCTATA
         GGTCTGAGGTAAGGCCCAAGAATTTGCATATTTGCATTTCGTTTTCTTTTCCTTTCTTTT
         [C,T]
         TTTTTTTTTTTTTTTGAGATGAAGTCTCACCCTGTCGCCCAGACTGGAGTGCAGTGGCAT
         GATCTCAGCTCACTGCAGCCTCTGCCTCCTGGGTTAAAGCGATTCTCCCCACACCCCAGA
         CCCGCTCCTGAGTAGCTGGGATTACAGGTGCCCGCCACCATGACTAGCTAACGTTTGTAT
         TTTTAGTAGAGACGGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCA
         GGTGATCCACTCACCTCAGCCTCCCAAGGTCTTGGGATTACTGGTGTGAGCCACCGCGTG

18074    TGCAGCCTCTGCCTCCTGGGTTAAAGCGATTCTCCCCACACCCCAGACCCGCTCCTGAGT
         AGCTGGGATTACAGGTGCCCGCCACCATGACTAGCTAACGTTTGTATTTTTAGTAGAGAC
         GGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGGTGATCCACTCA

FIGURE 3, page 12 of 14

```
        CCTCAGCCTCCCAAGGTCTTGGGATTACTGGTGTGAGCCACCGCGTGCGGCCAGAATTTG
        CATTTCTAACAAGTCCCAGGTGATGCTGATGCTGTGGGTCCAGGGACACACTTTGAGAAC
        [T,C,A]
        GCTTGTTACTCAGGCGATATGTGGACAGTAGCGTCATCTTCACCTGGGAGCTTCCTGCAG
        CATCTCAGGCCTTGCCCTACACCTACCAGATCAGAATCTGCATTTTAACTCAATCCCCGC
        GTGATTCTCATGCACCTGGAAGTTTGAGAAATATGACCTTAGAGGAGCCGGAATGTGAAA
        CCACTGGAGGCAGAGATAGATGGAGAATATCTCTTCTTCTCACGGATACTAAAGATGCAA
        CAAAAAGGGCTGACTCTCTGGGTGTGCACCCAGGTGGGGCTGATGACCGAAAAGAGGCCA

19328   TGTGTGTGAGGCCGGGGAGTGCTGCGAGCCCCGGAATTCCTCAGCCTTAGTCCCCCGCCA
        CATAGCTAAGAAGTGAGGGAGGAGGTGAGAAGGAGTCACTGCCCAGCCTCACTTCCGGTG
        GAGTACCCTGTCTCCTTGTCAGTTCTGTCTCTGGGGACAGTTGCCTGCTTTCACCTCTCC
        CTCCATCCCCTCTTCTCTCACAGGGAAAAATTCACCTTAATATTGGAAGTTCCTCTCCTA
        GCAAAGTCCTTCTCAGGCACCCACAGGCAAAAAGGAAACTAAGCAGAGTTAGGGCTTCCA
        [G,T]
        GCCTAGCCAACTACACGACTCTCCTCTTGCTTCCCTAAGAACCAGCGCAAGGGGCAGCGT
        GGGTTCCAGCATAGATGGACCTGTGTTGGAATCTCTGCACGTGCTGTGCTGACCCTGGCT
        AGCCATTGACCTCTCTGAGCCCTTGTTTCCTTTCCACTAGGCTCTCTGAGGGCAGGGGCC
        ATGTCTTTTTCACTGCTCTGTCTGCACTGAGCACTGTGCAGGGCACATAGGAAGTTCCCA
        TAAATGTTTGTGGGATAAAGGAAATAAAACCTTCTCTCTTCCTGTCCCCCTTGTGATGGC

19570   AAAGTCCTTCTCAGGCACCCACAGGCAAAAAGGAAACTAAGCAGAGTTAGGGCTTCCAGG
        CCTAGCCAACTACACGACTCTCCTCTTGCTTCCCTAAGAACCAGCGCAAGGGGCAGCGTG
        GGTTCCAGCATAGATGGACCTGTGTTGGAATCTCTGCACGTGCTGTGCTGACCCTGGCTA
        GCCATTGACCTCTCTGAGCCCTTGTTTCCTTTCCACTAGGCTCTCTGAGGGCAGGGGCCA
        TGTCTTTTTCACTGCTCTGTCTGCACTGAGCACTGTGCAGGGCACATAGGAAGTTCCCAT
        [A,G]
        AATGTTTGTGGGATAAAGGAAATAAAACCTTCTCTCTTCCTGTCCCCCTTGTGATGGCTT
        TGCACAAGGCACTGTCCTTGGCCAGGTTTGCTAGGCTAGTGTGAGGATAAACCAGGTATA
        TTACAAATTGGAGAAAATTTCTCGTTCTTCTTGGAAGAAGGTGCTGTATCATGAAACAAG
        AATGTCTTGATTCCCTTCTATGCCAGGTACTGGGGAGAAACAGGTGCCTGATAACCGTTG
        ATCCAGGCAGAAATAAGCATACTCCTGCTTCCCAAGGCCTGATGCTTCTCTCCTTCCTCC

20892   CCTTGGATGAAGAAGCGTGGGAACTCTTTGCTTCCTTTCCCTCCCGCAGTGACATGCCAT
        GCCATGCCACTGCCTCTTCATCTGGTCCTATGACAGTCACTCATAAGCACCCGCATGTAC
        CCGGCCCTGCACTAGCTCATGACAGCTGCAGTCAATTGGGCCAGGTGCTGTATCTCATCC
        GGCCTCCTCAGCAACCCTCTGAGATACTGGTAATGTCCCTGATGAAGATATTTACTGAGG
        CAGAAATGGACGCTCAGTGAAGCAAGGTGCCTGATGTTATAGCAATGAGCTATGAGTGGC
        [C,T]
        AGAGGGAGGAGATAAGCTCAGGCCTGACACCAAAGCCCATGCTCCTTCTAGTCAACCACA
        GTGCCTCCTATGGTGAATGAGTGAGTCAGCAACCAAGACGCATGAGGCCTTCTTTTTGGT
        GAGCCTTGGCTGGGTGCTGAGGCTTCAGGTACAATCATGGGTTGGAAGAGCCCTCCTCTC
        TCTCCACAGTCTGGCACTATGACCCCTTCTGGTTATTAACAAGGCAAAGAGAGAGAGGGA
        AGAAAGCAGGCAAATAATGTGGGTTGCTATTCCTAGAGATTAGAATTTCAGGAAGGATAA

26465   TTCTCTGACCCCTCCCCTCCGGTGCGTTTCGTATCAAAGCTCCTCAAACCCCGTCCCCCG
        TGTGTCCTGCTGTGTGCAGCTCGCTCTTTCCTTCCTTCCTAAGCTATCCAAGGGGATGGA
        CCCAGGCTCGTGGGGAGGTTCCACCCTTGGATCCAGGAAGAACCCTCCACCCTGCCTCGT
        GGGTGGGCCAAAGGCTACAGGGTGCTTCTTCCTCTTCCCCCACCCCCACTGTCCCTCATG
        TGCCATGGGCCTGCCTCCCCAGTGACCTGCGAAAGTGGAGCATCGAGGTAGGAGGGAAAC
        [G,A]
        GCAACCAGGGAGTCCTCGAGCCTGGGGCTGCCCTACCTCTACCCATTCCCCGACCAGAGC
        TTTGCCCTTGCTTGGCTGCCCGCCTGCCTCTTTGGGGAACTGAGCTCAGAGGCAGGTGCT
        TCAGAGAAGGAAACAAAATGAGGGGTGGCAGGGATAAAAAGTCACCTCCATTCTCTACCT
        CCCATGCAGCATGAACACAATTTCTCTCCACCTGGCTCCCAAATTTAAAGATGTGGACCA
        AGGCCTGTGGGTACTCCAGGGGCAAGGAGAGCCCTGGGGTCAGTGACACTGTCAGGCCAA

26472   ACCCCTCCCCTCCGGTGCGTTTCGTATCAAAGCTCCTCAAACCCCGTCCCCCGTGTGTCC
        TGCTGTGTGCAGCTCGCTCTTTCCTTCCTTCCTAAGCTATCCAAGGGGATGGACCCAGGC
        TCGTGGGGAGGTTCCACCCTTGGATCCAGGAAGAACCCTCCACCCTGCCTCGTGGGTGGG
        CCAAAGGCTACAGGGTGCTTCTTCCTCTTCCCCCACCCCCACTGTCCCTCATGTGCCATG
        GGCCTGCCTCCCCAGTGACCTGCGAAAGTGGAGCATCGAGGTAGGAGGGAAACGGCAACC
        [A,G]
```

# FIGURE 3, page 13 of 14

```
         GGGAGTCCTCGAGCCTGGGGCTGCCCTACCTCTACCCATTCCCCGACCAGAGCTTTGCCC
         TTGCTTGGCTGCCCGCCTGCCTCTTTGGGGAACTGAGCTCAGAGGCAGGTGCTTCAGAGA
         AGGAAACAAAATGAGGGGTGGCAGGGATAAAAAGTCACCTCCATTCTCTACCTCCCATGC
         AGCATGAACACAATTTCTCTCCACCTGGCTCCCAAATTTAAAGATGTGGACCAAGGCCTG
         TGGGTACTCCAGGGGCAAGGAGAGCCCTGGGGTCAGTGACACTGTCAGGCCAACCATGCA

28071    GCCAGGGACTTGAGAAGTTATATTGGGCAGTGGCTCCAATCTGTGGACCAGTATTTCAGC
         TTTCCCTGAAGATCAGGCAGGGTGCCATTCATTGTCTTTCTCTCCTAGCCCCCTCAGGAA
         AGAAGGACTATATTTGTACTGTACCCTAGGGGTTCTGGAAGGGAAAACATGGAATCAGGA
         TTCTATAGACTGATAGGCCCTATCCACAAGGGCCATGACTGGGAAAAGGTATGGGAGCAG
         AAGGAGAATTGGGATTTTAGGGTGCAGCTACGCTCACCCTAAACTTTTGGTGGCCTGGGG
         [C,T]
         ATGTCTTGAGGCCCAGACTGTTAACCAGGCTCTGCTGGCCTGTTTACTCGTCACCACCTC
         TGCACCTGCTGTCTTGAGACTCCATCCAGCCCCAGGCACGCCACCTGCTCCTGAGCCTCC
         ACTATCTCCCTGTGACGGGTGAACTTCGTGTACTGTGTCTCGGGTCCATATATGAATTGT
         GAGCAGGGTTCATCTATTTTAAACACAGATGTTTACAAAATAAAGATTATTTCAAACCAC
         CGGTGTGGCTGCCTGGATGAGTCCTTGGGGGTAGGTCTCACTCAGACCCTGGCAGTGATG

28096    GGCAGTGGCTCCAATCTGTGGACCAGTATTTCAGCTTTCCCTGAAGATCAGGCAGGGTGC
         CATTCATTGTCTTTCTCTCCTAGCCCCCTCAGGAAAGAAGGACTATATTTGTACTGTACC
         CTAGGGGTTCTGGAAGGGAAAACATGGAATCAGGATTCTATAGACTGATAGGCCCTATCC
         ACAAGGGCCATGACTGGGAAAAGGTATGGGAGCAGAAGGAGAATTGGGATTTTAGGGTGC
         AGCTACGCTCACCCTAAACTTTTGGTGGCCTGGGGCATGTCTTGAGGCCCAGACTGTTAA
         [C,G]
         CAGGCTCTGCTGGCCTGTTTACTCGTCACCACCTCTGCACCTGCTGTCTTGAGACTCCAT
         CCAGCCCCAGGCACGCCACCTGCTCCTGAGCCTCCACTATCTCCCTGTGACGGGTGAACT
         TCGTGTACTGTGTCTCGGGTCCATATATGAATTGTGAGCAGGGTTCATCTATTTTAAACA
         CAGATGTTTACAAAATAAAGATTATTTCAAACCACCGGTGTGGCTGCCTGGATGAGTCCT
         TGGGGGTAGGTCTCACTCAGACCCTGGCAGTGATGTGGGAGGGAGAGAGGCAGTGCTGGT

28403    CTGCTGGCCTGTTTACTCGTCACCACCTCTGCACCTGCTGTCTTGAGACTCCATCCAGCC
         CCAGGCACGCCACCTGCTCCTGAGCCTCCACTATCTCCCTGTGACGGGTGAACTTCGTGT
         ACTGTGTCTCGGGTCCATATATGAATTGTGAGCAGGGTTCATCTATTTTAAACACAGATG
         TTTACAAAATAAAGATTATTTCAAACCACCGGTGTGGCTGCCTGGATGAGTCCTTGGGGG
         TAGGTCTCACTCAGACCCTGGCAGTGATGTGGGAGGGAGAGAGGCAGTGCTGGTAGAAGC
         [A,G]
         GCTCCAGAAGCAAAGGCAACAGCAGTAGAGTGACCACGGAAGCGGCAAACATTGTCTTCC
         CTTCTCTACCTTCCCTAGTGCCACCTGCAGGGAGGCCCAAAGCAAAGCCCCGTTGCCCTG
         CATTGGGCTGGCACTGCAGAAATAAGATGAAACACAGTTATCGAGAGGATGCTGAACATC
         TATGAGCAGGTTTTAAAGCCAAGATGAGTCTCATCTGTTTGTGTGGGTCAGGAACGGGTC
         TTCCTGAAGGCATGAGGTGGGACTGGATAATCTTTCAGATTTGTGATTGGATACCTCGGG

28467    GCACGCCACCTGCTCCTGAGCCTCCACTATCTCCCTGTGACGGGTGAACTTCGTGTACTG
         TGTCTCGGGTCCATATATGAATTGTGAGCAGGGTTCATCTATTTTAAACACAGATGTTTA
         CAAAATAAAGATTATTTCAAACCACCGGTGTGGCTGCCTGGATGAGTCCTTGGGGGTAGG
         TCTCACTCAGACCCTGGCAGTGATGTGGGAGGGAGAGAGGCAGTGCTGGTAGAAGCAGCT
         CCAGAAGCAAAGGCAACAGCAGTAGAGTGACCACGGAAGCGGCAAACATTGTCTTCCCTT
         [C,G]
         TCTACCTTCCCTAGTGCCACCTGCAGGGAGGCCCAAAGCAAAGCCCCGTTGCCCTGCATT
         GGGCTGGCACTGCAGAAATAAGATGAAACACAGTTATCGAGAGGATGCTGAACATCTATG
         AGCAGGTTTTAAAGCCAAGATGAGTCTCATCTGTTTGTGTGGGTCAGGAACGGGTCTTCC
         TGAAGGCATGAGGTGGGACTGGATAATCTTTCAGATTTGTGATTGGATACCTCGGGGGAG
         CAGAGGCAGACTGGGATCTCAGGACTGCAGGTATTTCATACTTTGGGATATGGAATTGAT
```